Part III — Advanced Probability

Based on lectures by M. Lis

Notes taken by Dexter Chua

Michaelmas 2017

These notes are not endorsed by the lecturers, and I have modified them (often significantly) after lectures. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

The aim of the course is to introduce students to advanced topics in modern probability theory. The emphasis is on tools required in the rigorous analysis of stochastic processes, such as Brownian motion, and in applications where probability theory plays an important role.

Review of measure and integration: sigma-algebras, measures and filtrations; integrals and expectation; convergence theorems; product measures, independence and Fubini's theorem.
Conditional expectation: Discrete case, Gaussian case, conditional density functions; existence and uniqueness; basic properties.
Martingales: Martingales and submartingales in discrete time; optional stopping; Doob's inequalities, upcrossings, martingale convergence theorems; applications of martingale techniques.
Stochastic processes in continuous time: Kolmogorov's criterion, regularization of paths; martingales in continuous time.
Weak convergence: Definitions and characterizations; convergence in distribution, tightness, Prokhorov's theorem; characteristic functions, Lévy's continuity theorem.
Sums of independent random variables: Strong laws of large numbers; central limit theorem; Cramér's theory of large deviations.
Brownian motion: Wiener's existence theorem, scaling and symmetry properties; martingales associated with Brownian motion, the strong Markov property, hitting times; properties of sample paths, recurrence and transience; Brownian motion and the Dirichlet problem; Donsker's invariance principle.
Poisson random measures: Construction and properties; integrals.
Lévy processes: Lévy-Khinchin theorem.

Pre-requisites

A basic familiarity with measure theory and the measure-theoretic formulation of probability theory is very helpful. These foundational topics will be reviewed at the beginning of the course, but students unfamiliar with them are expected to consult the literature (for instance, Williams' book) to strengthen their understanding.

Contents

0 Introduction **3**

1 Some measure theory **4**
1.1 Review of measure theory 4
1.2 Conditional expectation 6

2 Martingales in discrete time **14**
2.1 Filtrations and martingales 14
2.2 Stopping time and optimal stopping 15
2.3 Martingale convergence theorems 18
2.4 Applications of martingales 24

3 Continuous time stochastic processes **29**

4 Weak convergence of measures **37**

5 Brownian motion **42**
5.1 Basic properties of Brownian motion 42
5.2 Harmonic functions and Brownian motion 48
5.3 Transience and recurrence 52
5.4 Donsker's invariance principle 53

6 Large deviations **57**

Index **60**

0 Introduction

In some other places in the world, this course might be known as "Stochastic Processes". In addition to doing probability, a new component studied in the course is *time*. We are going to study how things change over time.

In the first half of the course, we will focus on discrete time. A familiar example is the simple random walk — we start at a point on a grid, and at each time step, we jump to a neighbouring grid point randomly. This gives a sequence of random variables indexed by discrete time steps, and are related to each other in interesting ways. In particular, we will consider *martingales*, which enjoy some really nice convergence and "stopping "properties.

In the second half of the course, we will look at continuous time. There is a fundamental difference between the two, in that there is a nice topology on the interval. This allows us to say things like we want our trajectories to be continuous. On the other hand, this can cause some headaches because $\mathbb{R}$ is uncountable. We will spend a lot of time thinking about Brownian motion, whose discovery is often attributed to Robert Brown. We can think of this as the limit as we take finer and finer steps in a random walk. It turns out this has a very rich structure, and will tell us something about Laplace's equation as well.

Apart from stochastic processes themselves, there are two main objects that appear in this course. The first is the conditional expectation. Recall that if we have a random variable X, we can obtain a number $\mathbb{E}[X]$, the expectation of X. We can think of this as integrating out all the randomness of the system, and just remembering the average. Conditional expectation will be some subtle modification of this construction, where we don't actually get a number, but another random variable. The idea behind this is that we want to integrate out some of the randomness in our random variable, but keep the remaining.

Another main object is *stopping time*. For example, if we have a production line that produces random number of outputs at each point, then we can ask how much time it takes to produce a fixed number of goods. This is a nice random time, which we call a stopping time. The niceness follows from the fact that when the time comes, we know it. An example that is not nice is, for example, the last day it rains in Cambridge in a particular month, since on that last day, we don't necessarily know that it is in fact the last day.

At the end of the course, we will say a little bit about large deviations.

1 Some measure theory

1.1 Review of measure theory

To make the course as self-contained as possible, we shall begin with some review of measure theory. On the other hand, if one doesn't already know measure theory, they are recommended to learn the measure theory properly before starting this course.

Definition (σ-algebra)**.** Let E be a set. A subset $\mathcal{E}$ of the power set $\mathcal{P}(E)$ is called a *σ-algebra* (or *σ-field*) if

(i) $\emptyset \in \mathcal{E}$;

(ii) If $A \in \mathcal{E}$, then $A^C = E \setminus A \in \mathcal{E}$;

(iii) If $A_1, A_2, \ldots \in \mathcal{E}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{E}$.

Definition (Measurable space)**.** A *measurable space* is a set with a σ-algebra.

Definition (Borel σ-algebra)**.** Let E be a topological space with topology $\mathcal{T}$. Then the *Borel σ-algebra* $\mathcal{B}(E)$ on E is the σ-algebra generated by $\mathcal{T}$, i.e. the smallest σ-algebra containing $\mathcal{T}$.

We are often going to look at $\mathcal{B}(\mathbb{R})$, and we will just write $\mathcal{B}$ for it.

Definition (Measure)**.** A function $\mu : \mathcal{E} \to [0, \infty]$ is a *measure* if

- $\mu(\emptyset) = 0$
- If $A_1, A_2, \ldots \in \mathcal{E}$ are disjoint, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Definition (Measure space)**.** A *measure space* is a measurable space with a measure.

Definition (Measurable function)**.** Let $(E_1, \mathcal{E}_1)$ and $(E_2, \mathcal{E}_2)$ be measurable spaces. Then $f : E_1 \to E_2$ is said to be *measurable* if $A \in \mathcal{E}_2$ implies $f^{-1}(A) \in \mathcal{E}_1$.

This is similar to the definition of a continuous function.

Notation. For $(E, \mathcal{E})$ a measurable space, we write $m\mathcal{E}$ for the set of measurable functions $E \to \mathbb{R}$.

We write $m\mathcal{E}^+$ to be the positive measurable functions, which are allowed to take value ∞.

Note that we do *not* allow taking the values $\pm\infty$ in the first case.

Theorem. Let $(E, \mathcal{E}, \mu)$ be a measure space. Then there exists a unique function $\tilde{\mu} : m\mathcal{E}^+ \to [0, \infty]$ satisfying

- $\tilde{\mu}(\mathbf{1}_A) = \mu(A)$, where $\mathbf{1}_A$ is the indicator function of A.
- Linearity: $\tilde{\mu}(\alpha f + \beta g) = \alpha\tilde{\mu}(f) + \beta\tilde{\mu}(g)$ if $\alpha, \beta \in \mathbb{R}_{\geq 0}$ and $f, g \in m\mathcal{E}^+$.

– Monotone convergence: iff $f_1, f_2, \ldots \in m\mathcal{E}^+$ are such that $f_n \nearrow f \in m\mathcal{E}^+$ pointwise a.e. as $n \to \infty$, then

$$\lim_{n\to\infty} \tilde{\mu}(f_n) = \tilde{\mu}(f).$$

We call $\tilde{\mu}$ the *integral* with respect to μ, and we will write it as μ from now on.

Definition (Simple function)**.** A function f is *simple* if there exists $\alpha_n \in \mathbb{R}_{\geq 0}$ and $A_n \in \mathcal{E}$ for $1 \leq n \leq k$ such that

$$f = \sum_{n=1}^{k} \alpha_n \mathbf{1}_{A_n}.$$

From the first two properties of the measure, we see that

$$\mu(f) = \sum_{n=1}^{k} \alpha_n \mu(A_n).$$

One convenient observation is that a function is simple iff it takes on only finitely many values. We then see that if $f \in m\mathcal{E}^+$, then

$$f_n = 2^{-n}\lfloor 2^n f \rfloor \wedge n$$

is a sequence of simple functions approximating f from below. Thus, given monotone convergence, this shows that

$$\mu(f) = \lim \mu(f_n),$$

and this proves the uniqueness part of the theorem.

Recall that

Definition (Almost everywhere)**.** We say $f = g$ almost everywhere if

$$\mu(\{x \in E : f(x) \neq g(x)\}) = 0.$$

We say f is a *version* of g.

Example. Let $\ell_n = \mathbf{1}_{[n,n+1]}$. Then $\mu(\ell_n) = 1$ for all 1, but also $f_n \to 0$ and $\mu(0) = 0$. So the "monotone" part of monotone convergence is important.

So if the sequence is not monotone, then the measure does not preserve limits, but it turns out we still have an inequality.

Lemma (Fatou's lemma)**.** Let $f_i \in m\mathcal{E}^+$. Then

$$\mu\left(\liminf_n f_n\right) \leq \liminf_n \mu(f_n).$$

Proof. Apply monotone convergence to the sequence $\inf_{m\geq n} f_m$ □

Of course, it would be useful to extend integration to functions that are not necessarily positive.

Definition (Integrable function)**.** We say a function $f \in m\mathcal{E}$ is *integrable* if $\mu(|f|) \leq \infty$. We write $L^1(E)$ (or just L^1) for the space of integrable functions.

We extend μ to L^1 by

$$\mu(f) = \mu(f^+) - \mu(f^-),$$

where $f^\pm = (\pm f) \wedge 0$.

If we want to be explicit about the measure and the σ-algebra, we can write $L^1(E, \mathcal{E}\mu)$.

Theorem (Dominated convergence theorem)**.** If $f_i \in m\mathcal{E}$ and $f_i \to f$ a.e., such that there exists $g \in L^1$ such that $|f_i| \leq g$ a.e. Then

$$\mu(f) = \lim \mu(f_n).$$

Proof. Apply Fatou's lemma to $g - f_n$ and $g + f_n$. □

Definition (Product σ-algebra)**.** Let $(E_1, \mathcal{E}_1)$ and $(E_2, \mathcal{E}_2)$ be measure spaces. Then the product σ-algebra$\mathcal{E}_1 \otimes \mathcal{E}_2$ is the smallest σ-algebra on $E_1 \times E_2$ containing all sets of the form $A_1 \times A_2$, where $A_i \in \mathcal{E}_i$.

Theorem. If $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ are σ-finite measure spaces, then there exists a unique measure μ on $\mathcal{E}_1 \otimes \mathcal{E}_2)$ satisfying

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

for all $A_i \in \mathcal{E}_i$.

This is called the *product measure*.

Theorem (Fubini's/Tonelli's theorem)**.** If $f = f(x_1, x_2) \in m\mathcal{E}^+$ with $\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2$, then the functions

$$x_1 \mapsto \int f(x_1, x_2) \mathrm{d}\mu_2(x_2) \in m\mathcal{E}_1^+$$
$$x_2 \mapsto \int f(x_1, x_2) \mathrm{d}\mu_1(x_1) \in m\mathcal{E}_2^+$$

and

$$\begin{aligned}\int_E f \, \mathrm{d}u &= \int_{E_1} \left(\int_{E_2} f(x_1, x_2) \, \mathrm{d}\mu_2(x_2) \right) \mathrm{d}\mu_1(x_1) \\ &= \int_{E_2} \left(\int_{E_1} f(x_1, x_2) \, \mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2)\end{aligned}$$

1.2 Conditional expectation

In this course, conditional expectation is going to play an important role, and it is worth spending some time developing the theory. We are going to focus on probability theory, which, mathematically, just means we assume $\mu(E) = 1$. Practically, it is common to change notation to $E = \Omega$, $\mathcal{E} = \mathcal{F}$, $\mu = \mathbb{P}$ and $\int \mathrm{d}\mu = \mathbb{E}$. Measurable functions will be written as X, Y, Z, and will be called *random variables*. Elements in $\mathcal{F}$ will be called *events*. An element $\omega \in \Omega$ will be called a *realization*.

There are many ways we can think about conditional expectations. The first one is how most of us first encountered conditional probability.

Suppose $B \in \mathcal{F}$, with $\mathbb{P}(B) > 0$. Then the conditional probability of the event A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

This should be interpreted as the probability that A happened, given that B happened. Since we assume B happened, we ought to restrict to the subset of the probability space where B in fact happened. To make this a probability space, we scale the probability measure by $\mathbb{P}(B)$. Then given any event A, we take the probability of $A \cap B$ under this probability measure, which is the formula given.

More generally, if X is a random variable, the conditional expectation of f given B is just the expectation under this new probability measure,

$$\mathbb{E}[X \mid B] = \frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}[B]}.$$

We probably already know this from high school, and we are probably not quite excited by this. One natural generalization would be to allow B to vary.

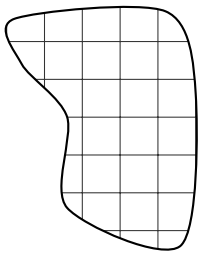
Let $G_1, G_2, \ldots \in \mathcal{F}$ be disjoint events such that $\bigcup_n G_n = \Omega$. Let

$$\mathcal{G} = \sigma(G_1, G_2, \ldots) = \left\{ \bigcup_{n \in I} G_n : I \subseteq \mathbb{N} \right\}.$$

Let $X \in L^1$. We then define

$$Y = \sum_{n=1}^{\infty} \mathbb{E}(X \mid G_n)\mathbf{1}_{G_n}.$$

Let's think about what this is saying. Suppose a random outcome ω happens. To compute Y, we figure out which of the G_n our ω belongs to. Let's say $\omega \in G_k$. Then Y returns the expected value of X given that we live in G_k. In this processes, we have forgotten the exact value of ω. All that matters is which G_n the outcome belongs to. We can "visually" think of the G_n as cutting up the sample space Ω into compartments:



We then average out X in each of these compartments to obtain Y. This is what we are going to call the conditional expectation of X given $\mathcal{G}$, written $\mathbb{E}(X \mid \mathcal{G})$.

Ultimately, the characterizing property of Y is the following lemma:

Lemma. The conditional expectation $Y = \mathbb{E}(X \mid \mathcal{G})$ satisfies the following properties:

- Y is $\mathcal{G}$-measurable
- We have $Y \in L^1$, and
$$\mathbb{E}Y\mathbf{1}_A = \mathbb{E}X\mathbf{1}_A$$
for all $A \in \mathcal{G}$.

Proof. It is clear that Y is $\mathcal{G}$-measurable. To show it is L^1, we compute

$$\begin{aligned}
\mathbb{E}[|Y|] &= \mathbb{E}\left|\sum_{n=1}^{\infty} \mathbb{E}(X \mid G_n)\mathbf{1}_{G_n}\right| \\
&\leq \mathbb{E}\sum_{n=1}^{\infty} \mathbb{E}(|X| \mid G_n)\mathbf{1}_{G_n} \\
&= \sum \mathbb{E}\left(\mathbb{E}(|X| \mid G_n)\mathbf{1}_{G_n}\right) \\
&= \sum \mathbb{E}|X|\mathbf{1}_{G_n} \\
&= \mathbb{E}\sum |X|\mathbf{1}_{G_n} \\
&= \mathbb{E}|X| \\
&< \infty,
\end{aligned}$$

where we used monotone convergence twice to swap the expectation and the sum.

The final part is also clear, since we can explicitly enumerate the elements in $\mathcal{G}$ and see that they all satisfy the last property. □

It turns out for any σ-subalgebra $\mathcal{G} \subseteq \mathcal{F}$, we can construct the conditional expectation $\mathbb{E}(X \mid \mathcal{G})$, which is uniquely characterized by the above two properties.

Theorem (Existence and uniqueness of conditional expectation)**.** Let $X \in L^1$, and $\mathcal{G} \subseteq \mathcal{F}$. Then there exists a random variable Y such that

- Y is $\mathcal{G}$-measurable
- $Y \in L^1$, and $\mathbb{E}X\mathbf{1}_A = \mathbb{E}Y\mathbf{1}_A$ for all $A \in \mathcal{G}$.

Moreover, if Y' is another random variable satisfying these conditions, then $Y' = Y$ almost surely.

We call Y a (version of) the conditional expectation given $\mathcal{G}$.

We will write the condition expectation as $\mathbb{E}(X \mid \mathcal{G})$, and if $X = \mathbf{1}_A$, we will write $\mathbb{E}(A \mid \mathcal{G}) = \mathbb{E}(\mathbf{1}_A \mid \mathcal{G})$.

Recall also that if Z is a random variable, then $\sigma(Z) = \{Z^{-1}(B) : B \in \mathcal{B}\}$. In this case, we will write $\mathbb{E}(X \mid Z) = \mathbb{E}(X \mid \sigma(Z))$.

By, say, bounded convergence, it follows from the second condition that $\mathbb{E}XZ = \mathbb{E}YZ$ for all bounded $\mathcal{G}$-measurable functions Z.

Proof. We first consider the case where $X \in L^2(\Omega, \mathcal{F}, \mu)$. Then we know from functional analysis that for any $\mathcal{G} \subseteq \mathcal{F}$, the space $L^2(\mathcal{G})$ is a Hilbert space with inner product

$$\langle X, Y\rangle = \mu(XY).$$

In particular, $L^2(\mathcal{G})$ is a closed subspace of $L^2(\mathcal{F})$. We can then define Y to be the orthogonal projection of X onto $L^2(\mathcal{G})$. It is immediate that Y is $\mathcal{G}$-measurable. For the second part, we use that $X - Y$ is orthogonal to $L^2(\mathcal{G})$, since that's what orthogonal projection is supposed to be. So $\mathbb{E}(X - Y)Z = 0$ for all $Z \in L^2(\mathcal{G})$. In particular, since the measure space is finite, the indicator function of any measurable subset is L^2. So we are done.

We next focus on the case where $X \in m\mathcal{E}^+$. We define

$$X_n = X \wedge n$$

We want to use monotone convergence to obtain our result. To do so, we need the following result:

Claim. If (X, Y) and (X', Y') satisfy the conditions of the theorem, and $X' \geq X$ a.s., then $Y' \geq Y$ a.s.

Proof. Define the event $A = \{Y' \leq Y\} \in \mathcal{G}$. Consider the event $Z = (Y - Y')\mathbf{1}_A$. Then $Z \geq 0$. We then have

$$\mathbb{E}Y'\mathbf{1}_A = \mathbb{E}X'\mathbf{1}_A \geq \mathbb{E}X\mathbf{1}_A = \mathbb{E}Y\mathbf{1}_A.$$

So it follows that we also have $\mathbb{E}(Y - Y')\mathbf{1}_A \leq 0$. So in fact $\mathbb{E}Z = 0$. So $Y' \geq Y$ a.s. □

We can now define $Y_n = \mathbb{E}(X_n \mid \mathcal{G})$, picking them so that $\{Y_n\}$ is increasing. We then take $Y_\infty = \lim Y_n$. Then Y_∞ is certainly $\mathcal{G}$-measurable, and by monotone convergence, if $A \in \mathcal{G}$, then

$$\mathbb{E}X\mathbf{1}_A = \lim \mathbb{E}X_n\mathbf{1}_A = \lim \mathbb{E}Y_n\mathbf{1}_A = \mathbb{E}Y_\infty\mathbf{1}_A.$$

Now if $\mathbb{E}X < \infty$, then $\mathbb{E}Y_\infty = \mathbb{E}X < \infty$. So we know Y_∞ is finite a.s., and we can define $Y = Y_\infty\mathbf{1}_{Y_\infty < \infty}$.

Finally, we work with arbitrary $X \in L^1$. We can write $X = X^+ - X^-$, and then define $Y^\pm = \mathbb{E}(X^\pm \mid \mathcal{G})$, and take $Y = Y^+ - Y^-$.

Uniqueness is then clear. □

Lemma. If Y is $\sigma(Z)$-measurable, then there exists $h : \mathbb{R} \to \mathbb{R}$ Borel-measurable such that $Y = h(Z)$. In particular,

$$\mathbb{E}(X \mid Z) = h(Z) \text{ a.s.}$$

for some $h : \mathbb{R} \to \mathbb{R}$.

We can then define $\mathbb{E}(X \mid Z = z) = h(z)$. The point of doing this is that we want to allow for the case where in fact we have $\mathbb{P}(Z = z) = 0$, in which case our original definition does not make sense.

Exercise. Consider $X \in L^1$, and $Z : \Omega \to \mathbb{N}$ discrete. Compute $\mathbb{E}(X \mid Z)$ and compare our different definitions of conditional expectation.

Example. Let $(U, V) \in \mathbb{R}^2$ with density $f_{U,V}(u, v)$, so that for any $B_1, B_2 \in \mathcal{B}$, we have

$$\mathbb{P}(U \in B_1, V \in B_2) = \int_{B_1}\int_{b_2} f_{U,V}(u, v)\,\mathrm{d}u\,\mathrm{d}v.$$

We want to compute $\mathbb{E}(h(V) \mid U)$, where $h : \mathbb{R} \to \mathbb{R}$ is Borel measurable. We can define

$$f_U(u) = \int_{\mathbb{R}} f_{U,V}(u, v)\, \mathrm{d}v,$$

and we define the conditional density of V given U by

$$F_{|U}(v \mid u) = \frac{f_{U,V}(u, v)}{f_U(u)}.$$

We define

$$g(u) = \int h(u) f_{V|U}(v \mid u)\, \mathrm{d}v.$$

We claim that $\mathbb{E}(h(V) \mid U)$ is just $g(U)$.

To check this, we show that it satisfies the two desired conditions. It is clear that it is $\sigma(U)$-measurable. To check the second condition, fix an $A \in \sigma(U)$. Then $A = \{(u, v) : u \in B\}$ for some B. Then

$$\begin{aligned}
\mathbb{E}(h(V)\mathbf{1}_A) &= \iint h(v)\mathbf{1}_{u\in B} f_{U,V}(u, v)\, \mathrm{d}u\, \mathrm{d}v \\
&= \iint h(v)\mathbf{1}_{u\in B} f_{V|U}(v \mid u) f_V(u)\, \mathrm{d}u\, \mathrm{d}v \\
&= \int g(U)\mathbf{1}_{u\in B} f_U(u)\, \mathrm{d}u \\
&= \mathbb{E}(g(U)\mathbf{1}_A),
\end{aligned}$$

as desired.

The point of this example is that to compute conditional expectations, we use our intuition to guess what the conditional expectation should be, and then check that it satisfies the two uniquely characterizing properties.

Example. Suppose (X, W) are Gaussian. Then for all linear functions $\varphi : \mathbb{R}^2 \to \mathbb{R}$, the quantity $\varphi(X, W)$ is Gaussian.

One nice property of Gaussians is that lack of correlation implies independence. We want to compute $\mathbb{E}(X \mid W)$. Note that if Y is such that $\mathbb{E}X = \mathbb{E}Y$, $X - Y$ is independent of W, and Y is W-measurable, then $Y = \mathbb{E}(X \mid W)$, since $\mathbb{E}(X - Y)\mathbf{1}_A = 0$ for all $\sigma(W)$-measurable A.

The guess is that we want Y to be a Gaussian variable. We put $Y = aW + b$. Then $\mathbb{E}X = \mathbb{E}Y$ implies we must have

$$a\mathbb{E}W + b = \mathbb{E}X. \tag{$*$}$$

The independence part requires $\operatorname{cov}(X - Y, W) = 0$. Since covariance is linear, we have

$$0 = \operatorname{cov}(X - Y, W) = \operatorname{cov}(X, W) - \operatorname{cov}(aW + b, W) = \operatorname{cov}(X, W) - a \operatorname{cov}(W, W).$$

Recalling that $\operatorname{cov}(W, W) = \operatorname{var}(W)$, we need

$$a = \frac{\operatorname{cov}(X, W)}{\operatorname{var}(W)}.$$

This then allows us to use $(*)$ to compute b as well. This is how we compute the conditional expectation of Gaussians.

We note some immediate properties of conditional expectation. As usual, all (in)equality and convergence statements are to be taken with the quantifier "almost surely".

Proposition.

(i) $\mathbb{E}(X \mid \mathcal{G}) = X$ iff X is $\mathcal{G}$-measurable.

(ii) $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}X$

(iii) If $X \geq 0$ a.s., then $\mathbb{E}(X \mid \mathcal{G}) \geq 0$

(iv) If X and $\mathcal{G}$ are independent, then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}[X]$

(v) If $\alpha, \beta \in \mathbb{R}$ and $X_1, X_2 \in L^1$, then

$$\mathbb{E}(\alpha X_1 + \beta X_2 \mid \mathcal{G}) = \alpha \mathbb{E}(X_1 \mid \mathcal{G}) + \beta \mathbb{E}(X_2 \mid \mathcal{G}).$$

(vi) Suppose $X_n \nearrow X$. Then

$$\mathbb{E}(X_n \mid \mathcal{G}) \nearrow \mathbb{E}(X \mid \mathcal{G}).$$

(vii) *Fatou's lemma*: If X_n are non-negative measurable, then

$$\mathbb{E}\left(\liminf_{n\to\infty} X_n \mid \mathcal{G}\right) \leq \liminf_{n\to\infty} \mathbb{E}(X_n \mid \mathcal{G}).$$

(viii) *Dominated convergence theorem*: If $X_n \to X$ and $Y \in L^1$ such that $Y \geq |X_n|$ for all n, then

$$\mathbb{E}(X_n \mid \mathcal{G}) \to \mathbb{E}(X \mid \mathcal{G}).$$

(ix) *Jensen's inequality*: If $c : \mathbb{R} \to \mathbb{R}$ is convex, then

$$\mathbb{E}(c(X) \mid \mathcal{G}) \geq c(\mathbb{E}(X) \mid \mathcal{G}).$$

(x) *Tower property*: If $\mathcal{H} \subseteq \mathcal{G}$, then

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H}).$$

(xi) For $p \geq 1$,

$$\|\mathbb{E}(X \mid \mathcal{G})\|_p \leq \|X\|_p.$$

(xii) If Z is bounded and $\mathcal{G}$-measurable, then

$$\mathbb{E}(ZX \mid \mathcal{G}) = Z\mathbb{E}(X \mid \mathcal{G}).$$

(xiii) Let $X \in L^1$ and $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$. Assume that $\sigma(X, \mathcal{G})$ is independent of $\mathcal{H}$. Then

$$\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X \mid \sigma(\mathcal{G}, \mathcal{H})).$$

Proof.

(i) Clear.

(ii) Take $A = \omega$.

(iii) Shown in the proof.

(iv) Clear by property of expected value of independent variables.

(v) Clear, since the RHS satisfies the unique characterizing property of the LHS.

(vi) Clear from construction.

(vii) Same as the unconditional proof, using the previous property.

(viii) Same as the unconditional proof, using the previous property.

(ix) Same as the unconditional proof.

(x) The LHS satisfies the characterizing property of the RHS

(xi) Using the convexity of $|x|^p$, Jensen's inequality tells us

$$\begin{aligned}\|E(X \mid \mathcal{G})\|_p^p &= \mathbb{E}|\mathbb{E}(X \mid \mathcal{G})|^p \\ &\leq \mathbb{E}(\mathbb{E}(|X|^p \mid \mathcal{G})) \\ &= \mathbb{E}|X|^p \\ &= \|X\|_p^p\end{aligned}$$

(xii) If $Z = \mathbf{1}_B$, and let $b \in \mathcal{G}$. Then

$$\mathbb{E}(Z\mathbb{E}(X \mid \mathcal{G})\mathbf{1}_A) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \cdot \mathbf{1}_{A\cap B}) = \mathbb{E}(X\mathbf{1}_{A\cap B}) = \mathbb{E}(ZX\mathbf{1}_A).$$

So the lemma holds. Linearity then implies the result for Z simple, then apply our favorite convergence theorems.

(xiii) Take $B \in \mathcal{H}$ and $A \in \mathcal{G}$. Then

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X \mid \sigma(\mathcal{G}, \mathcal{H})) \cdot \mathbf{1}_{A\cap B}) &= \mathbb{E}(X \cdot \mathbf{1}_{A\cap B}) \\ &= \mathbb{E}(X\mathbf{1}_A)\mathbb{P}(B) \\ &= \mathbb{E}(\mathbb{E}(X \mid \mathcal{G})\mathbf{1}_A)\mathbb{P}(B) \\ &= \mathbb{E}(\mathbb{E}(X \mid \mathcal{G})\mathbf{1}_{A\cap B})\end{aligned}$$

If instead of $A \cap B$, we had any $\sigma(\mathcal{G}, \mathcal{H})$-measurable set, then we would be done. But we are fine, since the set of subsets of the form $A \cap B$ with $A \in \mathcal{G}$, $B \in \mathcal{H}$ is a generating π-system for $\sigma(\mathcal{H}, \mathcal{G})$. □

We shall end with the following key lemma. We will later use it to show that many of our *martingales* are uniformly integrable.

Lemma. If $X \in L^1$, then the family of random variables $Y_\mathcal{G} = \mathbb{E}(X \mid \mathcal{G})$ for all $\mathcal{G} \subseteq \mathcal{F}$ is uniformly integrable.

In other words, for all $\varepsilon > 0$, there exists $\lambda > 0$ such that

$$\mathbb{E}(Y_\mathcal{G}\mathbf{1}_{|Y_\mathcal{G}>\lambda|}) < \varepsilon$$

for all $\mathcal{G}$.

Proof. Fix $\varepsilon > 0$. Then there exists $\delta > 0$ such that $\mathbb{E}|X|\mathbf{1}_A < \varepsilon$ for any A with $\mathbb{P}(A) < \delta$.

Take $Y = \mathbb{E}(X \mid \mathcal{G})$. Then by Jensen, we know

$$|Y| \leq \mathbb{E}(|X| \mid \mathcal{G})$$

In particular, we have

$$\mathbb{E}|Y| \leq \mathbb{E}|X|.$$

By Markov's inequality, we have

$$\mathbb{P}(|Y| \geq \lambda) \leq \frac{\mathbb{E}|Y|}{\lambda} \leq \frac{\mathbb{E}|X|}{\lambda}.$$

So take λ such that $\frac{\mathbb{E}|X|}{\lambda} < \delta$. So we have

$$\mathbb{E}(|Y|\mathbf{1}_{|Y|\geq\lambda}) \leq \mathbb{E}(\mathbb{E}(|X| \mid \mathcal{G})\mathbf{1}_{|Y|\geq\lambda}) = \mathbb{E}(|X|\mathbf{1}_{|Y|\geq\lambda}) < \varepsilon$$

using that $\mathbf{1}_{|Y|\geq\lambda}$ is a $\mathcal{G}$-measurable function. □

2 Martingales in discrete time

2.1 Filtrations and martingales

We would like to model some random variable that "evolves with time". For example, in a simple random walk, X_n could be the position we are at time n. To do so, we would like to have some σ-algebras $\mathcal{F}_n$ that tells us the "information we have at time n". This structure is known as a *filtration*.

Definition (Filtration)**.** A *filtration* is a sequence of σ-algebras $(\mathcal{F}_n)_{n\geq 0}$ such that $\mathcal{F} \supseteq \mathcal{F}_{n+1} \supseteq \mathcal{F}_n$ for all n. We define $\mathcal{F}_\infty = \sigma(\mathcal{F}_0, \mathcal{F}_1, \ldots) \subseteq \mathcal{F}$.

We will from now on assume $(\Omega, \mathcal{F}, \mathbb{P})$ is equipped with a filtration $(\mathcal{F}_n)_{n\geq 0}$.

Definition (Stochastic process in discrete time)**.** A *stochastic process* (in discrete time) is a sequence of random variables $(X_n)_{n\geq 0}$.

This is a very general definition, and in most cases, we would want X_n to interact nicely with our filtration.

Definition (Natural filtration)**.** The *natural filtration* of $(X_n)_{n\geq 0}$ is given by

$$\mathcal{F}_n^X = \sigma(X_1, \ldots, X_n).$$

Definition (Adapted process)**.** We say that $(X_n)_{n\geq 0}$ is *adapted* (to $(\mathcal{F}_n)_{n\geq 0}$) if X_n is $\mathcal{F}_n$-measurable for all $n \geq 0$. Equivalently, if $\mathcal{F}_n^X \subseteq \mathcal{F}_n$.

Definition (Integrable process)**.** A process $(X_n)_{n\geq 0}$ is *integrable* if $X_n \in L^1$ for all $n \geq 0$.

We can now write down the definition of a martingale.

Definition (Martingale)**.** An integrable adapted process $(X_n)_{n\geq 0}$ is a *martingale* if for all $n \geq m$, we have

$$\mathbb{E}(X_n \mid \mathcal{F}_m) = X_m.$$

We say it is a *super-martingale* if

$$\mathbb{E}(X_n \mid \mathcal{F}_m) \leq X_m,$$

and a *sub-martingale* if

$$\mathbb{E}(X_n \mid \mathcal{F}_m) \geq X_m,$$

Note that it is enough to take $m = n - 1$ for all $n \geq 0$, using the tower property.

The idea of a martingale is that we cannot predict whether X_n will go up or go down in the future even if we have all the information up to the present. For example, if X_n denotes the wealth of a gambler in a gambling game, then in some sense $(X_n)_{n\geq 0}$ being a martingale means the game is "fair" (in the sense of a fair dice).

Note that $(X_n)_{n\geq 0}$ is a super-martingale iff $(-X_n)_{n\geq 0}$ is a sub-martingale, and if $(X_n)_{n\geq 0}$ is a martingale, then it is both a super-martingale and a sub-martingale. Often, what these extra notions buy us is that we can formulate our results for super-martingales (or sub-martingales), and then by applying the result to both $(X_n)_{n\geq 0}$ and $(-X_n)_{n\geq 0}$, we obtain the desired, stronger result for martingales.

2.2 Stopping time and optimal stopping

The optional stopping theorem says the definition of a martingale in fact implies an *a priori* much stronger property. To formulate the optional stopping theorem, we need the notion of a stopping time.

Definition (Stopping time)**.** A *stopping time* is a random variable $T : \Omega \to \mathbb{N}_{\geq 0} \cup \{\infty\}$ such that

$$\{T \leq n\} \in \mathcal{F}_n$$

for all $n \geq 0$.

This means that at time n, if we want to know if T has occurred, we can determine it using the information we have at time n.

Note that T is a stopping time iff $\{t = n\} \in \mathcal{F}_n$ for all n, since if T is a stopping time, then

$$\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\},$$

and $\{T \leq n-1\} \in \mathcal{F}_{n-1} \subseteq \mathcal{F}_n$. Conversely,

$$\{T \leq n\} = \bigcup_{k=1}^{n} \{T = k\} \in \mathcal{F}_n.$$

This will not be true in the continuous case.

Example. If $B \in \mathcal{B}(\mathbb{R})$, then we can define

$$T = \inf\{n : X_n \in B\}.$$

Then this is a stopping time.

On the other hand,

$$T = \sup\{n : X_n \in B\}$$

is not a stopping time (in general).

Given a stopping time, we can make the following definition:

Definition (X_T)**.** For a stopping time T, we define the random variable X_T by

$$X_T(\omega) = X_{T(\omega)}(\omega)$$

on $\{T < \infty\}$, and 0 otherwise.

Later, for suitable martingales, we will see that the limit $X_\infty = \lim_{n\to\infty} X_n$ makes sense. In that case, We define $X_T(\omega)$ to be $X_\infty(\omega)$ if $T = \infty$.

Similarly, we can define

Definition (Stopped process)**.** The *stopped process* is defined by

$$(X_n^T)_{n\geq 0} = (X_{T(\omega)\wedge n}(\omega))_{n\geq 0}.$$

This says we stop evolving the random variable X once T has occurred.

We would like to say that X_T is "$\mathcal{F}_T$-measurable", i.e. to compute X_T, we only need to know the information up to time T. After some thought, we see that the following is the correct definition of $\mathcal{F}_T$:

Definition ($\mathcal{F}_T$)**.** For a stopping time T, define

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq n\} \in \mathcal{F}_n\}.$$

This is easily seen to be a σ-algebra.

Example. If $T \equiv n$ is constant, then $\mathcal{F}_T = \mathcal{F}_n$.

There are some fairly immediate properties of these objects, whose proof is left as an exercise for the reader:

Proposition.

(i) If $T, S, (T_n)_{n\geq 0}$ are all stopping times, then

$$T \vee S, T \wedge S, \sup_n T_n, \inf T_n, \limsup T_n, \liminf T_n$$

are all stopping times.

(ii) $\mathcal{F}_T$ is a σ-algebra

(iii) If $S \leq T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$.

(iv) $X_T \mathbf{1}_{T<\infty}$ is $\mathcal{F}_T$-measurable.

(v) If (X_n) is an adapted process, then so is $(X_n^T)_{n\geq 0}$ for any stopping time T.

(vi) If (X_n) is an integrable process, then so is $(X_n^T)_{n\geq 0}$ for any stopping time T. □

We now come to the fundamental property of martingales.

Theorem (Optional stopping theorem)**.** Let $(X_n)_{n\geq 0}$ be a super-martingale and $S \leq T$ *bounded* stopping times. Then

$$\mathbb{E}X_T \leq \mathbb{E}X_S.$$

Proof. Follows from the next theorem. □

What does this theorem mean? If X is a martingale, then it is both a super-martingale and a sub-martingale. So we can apply this to both X and $-X$, and so we have

$$\mathbb{E}(X_T) = \mathbb{E}(X_S).$$

In particular, since 0 is a stopping time, we see that

$$\mathbb{E}X_T = \mathbb{E}X_0$$

for *any* bounded stopping time T.

Recall that martingales are supposed to model fair games. If we again think of X_n as the wealth at time n, and T as the time we stop gambling, then this says no matter how we choose T, as long as it is bounded, the expected wealth at the end is the same as what we started with.

Example. Consider the stopping time

$$T = \inf\{n : X_n = 1\},$$

and take X such that $\mathbb{E}X_0 = 0$. Then clearly we have $\mathbb{E}X_T = 1$. So this tells us T is not a bounded stopping time!

Theorem. The following are equivalent:

(i) $(X_n)_{n\geq 0}$ is a super-martingale.

(ii) For any bounded stopping times T and any stopping time S,

$$\mathbb{E}(X_T \mid \mathcal{F}_S) \leq X_{S\wedge T}.$$

(iii) (X_n^T) is a super-martingale for any stopping time T.

(iv) For bounded stopping times S, T such that $S \leq T$, we have

$$\mathbb{E}X_T \leq \mathbb{E}X_S.$$

In particular, (iv) implies (i).

Proof.

– (ii) $\Rightarrow$ (iii): Consider $(X_n^{T'})_{\geq 0}$ for a stopping time T'. To check if this is a super-martingale, we need to prove that whenever $m \leq n$,

$$\mathbb{E}(X_{n\wedge T'} \mid \mathcal{F}_m) \leq X_{m\wedge T'}.$$

But this follows from (ii) above by taking $S = m$ and $T = T' \wedge n$.

– (ii) $\Rightarrow$ (iv): Clear by the tower law.

– (iii) $\Rightarrow$ (i): Take $T = \infty$.

– (i) $\Rightarrow$ (ii): Assume $T \leq n$. Then

$$\begin{aligned} X_T &= X_{S\wedge T} + \sum_{S\leq k<T} (X_{k+1} - X_k) \\ &= X_{S\wedge T} + \sum_{k=0}^{n} (X_{k+1} - X_k)\mathbf{1}_{S\leq k<T} \qquad (*) \end{aligned}$$

Now note that $\{S \leq k < T\} = \{S \leq k\} \cap \{T \leq k\}^c \in \mathcal{F}_k$. Let $A \in \mathcal{F}_S$. Then $A \cap \{S \leq k\} \in \mathcal{F}_k$ by definition of $\mathcal{F}_S$. So $A \cap \{S \leq k < T\} \in \mathcal{F}_k$.

Apply $\mathbb{E}$ to to $(*) \times \mathbf{1}_A$. Then we have

$$\mathbb{E}(X_T\mathbf{1}_A) = \mathbb{E}(X_{S\wedge T}\mathbf{1}_A) + \sum_{k=0}^{n} \mathbb{E}(X_{k+1} - X_k)\mathbf{1}_{A\cap\{S\leq k<T\}}.$$

But for all k, we know

$$\mathbb{E}(X_{k+1} - X_k)\mathbf{1}_{A\cap\{S\leq k<T\}} \leq 0,$$

since X is a super-martingale. So it follows that for all $A \in \mathcal{F}_S$, we have

$$\mathbb{E}(X_T \cdot \mathbf{1}_A) \leq \mathbb{E}(X_{S\wedge T}\mathbf{1}_A).$$

But since $X_{S\wedge T}$ is $\mathcal{F}_{S\wedge T}$ measurable, it is in particular $\mathcal{F}_S$ measurable. So it follows that for all $A \in \mathcal{F}_S$, we have

$$\mathbb{E}(\mathbb{E}(X_T \mid \mathcal{F}_S)\mathbf{1}_A) \leq \mathbb{E}(X_{S\wedge T}\mathbf{1}_A).$$

So the result follows.

– (iv) $\Rightarrow$ (i): Fix $m \leq n$ and $A \in \mathcal{F}_m$. Take

$$T = m\mathbf{1}_A + \mathbf{1}_{A^C}.$$

One then manually checks that this is a stopping time. Now note that

$$X_T = X_m\mathbf{1}_A + X_n\mathbf{1}_{A^C}.$$

So we have

$$\begin{aligned} 0 &\geq \mathbb{E}(X_n) - \mathbb{E}(X_T)\\ &= \mathbb{E}(X_n) - \mathbb{E}(X_n\mathbf{1}_{A^c}) - \mathbb{E}(X_m\mathbf{1}_A)\\ &= \mathbb{E}(X_n\mathbf{1}_A) - \mathbb{E}(X_m\mathbf{1}_A). \end{aligned}$$

Then the same argument as before gives the result. □

2.3 Martingale convergence theorems

One particularly nice property of martingales is that they have nice convergence properties. We shall begin by proving a pointwise version of martingale convergence.

Theorem (Almost sure martingale convergence theorem)**.** Suppose $(X_n)_{n\geq 0}$ is a super-martingale that is bounded in L^1, i.e. $\sup_n \mathbb{E}|X_n| < \infty$. Then there exists an $\mathcal{F}_\infty$-measurable $X_\infty \in L^1$ such that

$$X_n \to X_\infty \text{ a.s. as } n \to \infty.$$

To begin, we need a convenient characterization of when a series converges.

Definition (Upcrossing)**.** Let (x_n) be a sequence and (a, b) an interval. An *upcrossing* of (a, b) by (x_n) is a sequence $j, j + 1, \ldots, k$ such that $x_j \leq a$ and $x_k \geq b$. We define

$$\begin{aligned} U_n[a, b, (x_n)] &= \text{number of disjoint upcrossings contained in } \{1, \ldots, n\}\\ U[a, b, (x_n)] &= \lim_{n\to\infty} U_n[a, b, x]. \end{aligned}$$

We can then make the following elementary observation:

Lemma. Let $(x_n)_{n\geq 0}$ be a sequence of numbers. Then x_n converges in $\mathbb{R}$ if and only if

(i) $\liminf |x_n| < \infty$.

(ii) For all $a, b \in \mathbb{Q}$ with $a < b$, we have $U[a, b, (x_n)] < \infty$.

For our martingales, since they are bounded in L^1, Fatou's lemma tells us $\mathbb{E} \liminf |X_n| < \infty$. So $\liminf |X_n| = 0$ almost surely. Thus, it remains to show that for any fixed $a < b \in \mathbb{Q}$, we have $\mathbb{P}(U[a, b, (X_n)] = \infty) = 0$. This is a consequence of Doob's upcrossing lemma.

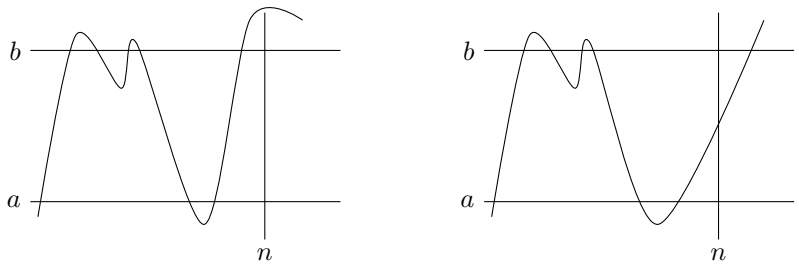
Lemma (Doob's upcrossing lemma)**.** If X_n is a super-martingale, then

$$(b - a)\mathbb{E}(U_n[a, b(X_n)]) \leq \mathbb{E}(X_n - a)^-$$

Proof. Assume that X is a positive super-martingale. We define stopping times S_k, T_k as follows:

- $T_0 = 0$
- $S_{k+1} = \inf\{n : X_n \leq a, n \geq T_n\}$
- $T_{k+1} = \inf\{n : X_n \geq b, n \geq S_{k+1}\}$.

Given an n, we want to count the number of upcrossings before n. There are two cases we might want to distinguish:



Now consider the sum

$$\sum_{k=1}^{n} X_{T_k \wedge n} - X_{S_k \wedge n}.$$

In the first case, this is equal to

$$\sum_{k=1}^{U_n} X_{T_k} - X_{S_k} + \sum_{k=U_n+1}^{n} X_n - X_n \geq (b - a)U_n.$$

In the second case, it is equal to

$$\sum_{k=1}^{U_n} X_{T_k} - X_{S_k} + (X_n - X_{S_{U_n+1}}) + \sum_{k=U_n+2}^{n} X_n - X_n \geq (b-a)U_n + (X_n - X_{S_{U_n+1}}).$$

Thus, in general, we have

$$\sum_{k=1}^{n} X_{T_k \wedge n} - X_{S_k \wedge n} \geq (b - a)U_n + (X_n - X_{S_{U_n+1} \wedge n}).$$

By definition, $S_k < T_k \leq n$. So the expectation of the LHS is always non-negative by super-martingale convergence, and thus

$$0 \geq (b-a)\mathbb{E}U_n + \mathbb{E}(X_n - X_{S_{U_n+1}\wedge n}).$$

Then observe that

$$X_n - X_{S_{U_n+1}} \geq -(X_n - a)^-. \qquad \square$$

The almost-sure martingale convergence theorem is very nice, but often it is not good enough. For example, we might want convergence in L^p instead. The following example shows this isn't always possible:

Example. Suppose $(\rho_n)_{n\geq 0}$ is a sequence of iid random variables and

$$\mathbb{P}(\rho_n = 0) = \frac{1}{2} = \mathbb{P}(\rho_n = 2).$$

Let

$$X_n = \prod_{k=0}^{n} \rho_k.$$

Then this is a martingale, and $\mathbb{E}X_n = 1$. On the other hand, $X_n \to 0$ almost surely. So $\|X_n - X_\infty\|_1$ does not converge to 0.

For $p > 1$, if we want convergence in L^p, it is not surprising that we at least need the sequence to be L^p bounded. We will see that this is in fact sufficient. For $p = 1$, however, we need a bit more than being bounded in L^1. We will need uniform integrability.

To prove this, we need to establish some inequalities.

Lemma (Maximal inequality)**.** Let (X_n) be a sub-martingale that is non-negative, or a martingale. Define

$$X_n^* = \sup_{k\leq n} |X_k|, \quad X^* = \lim_{n\to\infty} X_n^*.$$

If $\lambda \geq 0$, then

$$\lambda\mathbb{P}(X_n^* \geq \lambda) \leq \mathbb{E}[|X_n|\mathbf{1}_{X_n^*\geq\lambda}].$$

In particular, we have

$$\lambda\mathbb{P}(X_n^* \geq \lambda) \leq \mathbb{E}[|X_n|].$$

Markov's inequality says almost the same thing, but has $\mathbb{E}[|X_n^*|]$ instead of $\mathbb{E}[|X_n|]$. So this is a stronger inequality.

Proof. If X_n is a martingale, then $|X_n|$ is a sub-martingale. So it suffices to consider the case of a non-negative sub-martingale. We define the stopping time

$$T = \inf\{n : X_n \geq \lambda\}.$$

By optional stopping,

$$\begin{aligned}
\mathbb{E}X_n &\geq \mathbb{E}X_{T\wedge n} \\
&= \mathbb{E}X_T\mathbf{1}_{T\leq n} + \mathbb{E}X_n\mathbf{1}_{T>n} \\
&\geq \lambda\mathbb{P}(T \leq n) + \mathbb{E}X_n\mathbf{1}_{T>n} \\
&= \lambda\mathbb{P}(X_n^* \geq \lambda) + \mathbb{E}X_n\mathbf{1}_{T>n}.
\end{aligned}$$

Lemma (Doob's L^p inequality)**.** For $p > 1$, we have

$$\|X_n^*\|_p \leq \frac{p}{p-1}\|X_n\|_p$$

for all n.

Proof. Let $k > 0$, and consider

$$\|X_n^* \wedge k\|_p^p = \mathbb{E}|X_n^* \wedge k|^p.$$

We use the fact that

$$x^p = \int_0^x ps^{p-1}\,\mathrm{d}s.$$

So we have

$$\begin{aligned}
\|X_n^* \wedge k\|_p^p &= \mathbb{E}|X_n^* \wedge k|^p \\
&= \mathbb{E}\int_0^{X_n^* \wedge k} px^{p-1}\,\mathrm{d}x \\
&= \mathbb{E}\int_0^k px^{p-1}\mathbf{1}_{X_n^* \geq x}\,\mathrm{d}x \\
&= \int_0^k px^{p-1}\mathbb{P}(X_n^* \geq x)\,\mathrm{d}x && \text{(Fubini)} \\
&\leq \int_0^k px^{p-2}\mathbb{E}X_n\mathbf{1}_{X_n^* \geq x}\,\mathrm{d}x && \text{(maximal inequality)} \\
&= \mathbb{E}X_n\int_0^k px^{p-2}\mathbf{1}_{X_n^* \geq x}\,\mathrm{d}x && \text{(Fubini)} \\
&= \frac{p}{p-1}\mathbb{E}X_n(X_n^* \wedge k)^{p-1} \\
&\leq \frac{p}{p-1}\|X_n\|_p\left(\mathbb{E}(X_n^* \wedge k)^p\right)^{\frac{p-1}{p}} && \text{(H older)} \\
&= \frac{p}{p-1}\|X_n\|_p\|X_n^* \wedge k\|_p^{p-1}
\end{aligned}$$

Now take the limit $k \to \infty$ and divide by $\|X_n^*\|_p^{p-1}$. □

Theorem (L^p martingale convergence theorem)**.** Let $(X_n)_{n\geq 0}$ be a martingale, and $p > 1$. Then the following are equivalent:

(i) $(X_n)_{n\geq 0}$ is bounded in L^p, i.e. $\sup_n \mathbb{E}|X_i|^p < \infty$.

(ii) $(X_n)_{n\geq 0}$ converges as $n \to \infty$ to a random variable $X_\infty \in L^p$ almost surely and in L^p.

(iii) There exists a random variable $Z \in L^p$ such that

$$X_n = \mathbb{E}(Z \mid \mathcal{F}_n)$$

Moreover, in (iii), we always have $X_\infty = \mathbb{E}(Z \mid \mathcal{F}_\infty)$.

This gives a bijection between martingales bounded in L^p and $L^p(\mathcal{F}_\infty)$, sending $(X_n)_{n\geq 0} \mapsto X_\infty$.

Proof.

– (i) ⇒ (ii): If $(X_n)_{n\geq 0}$ is bounded in L^p, then it is bounded in L^1. So by the martingale convergence theorem, we know $(X_n)_{n\geq 0}$ converges almost surely to X_∞. By Fatou's lemma, we have $X_\infty \in L^p$.

Now by monotone convergence, we have

$$\|X^*\|_p = \lim_n \|X_n^*\|_p \leq \frac{p}{p-1} \sup_n \|X_n\|_p < \infty.$$

By the triangle inequality, we have

$$|X_n - X_\infty| \leq 2X^* \text{ a.s.}$$

So by dominated convergence, we know that $X_n \to X_\infty$ in L^p.

– (ii) ⇒ (iii): Take $Z = X_\infty$. We want to prove that

$$X_m = \mathbb{E}(X_\infty \mid \mathcal{F}_m).$$

To do so, we show that $\|X_m - \mathbb{E}(X_\infty \mid \mathcal{F}_m)\|_p = 0$. For $n \geq m$, we know this is equal to

$$\|E(X_n \mid \mathcal{F}_m) - \mathbb{E}(X_\infty \mid \mathcal{F}_m)\|_p = \|\mathbb{E}(X_n - X_\infty \mid \mathcal{F}_m)\|_p \leq \|X_n - X_\infty\|_p \to 0$$

as $n \to \infty$, where the last step uses Jensen's. But it is also a constant. So we are done.

– (iii) ⇒ (i): Since expectation decreases L^p norms, we already know that $(X_n)_{n\geq 0}$ is L^p-bounded.

To show the "moreover" part, note that $\bigcup_{n\geq 0} \mathcal{F}_n$ is a π-system that generates $\mathcal{F}_\infty$. So it is enough to prove that

$$\mathbb{E}X_\infty \mathbf{1}_A = \mathbb{E}(\mathbb{E}(Z \mid \mathcal{F}_\infty)\mathbf{1}_A).$$

But if $A \in \mathcal{F}_N$, then

$$\begin{aligned}\mathbb{E}X_\infty \mathbf{1}_A &= \lim_{n\to\infty} \mathbb{E}X_n \mathbf{1}_A \\ &= \lim_{n\to\infty} \mathbb{E}(\mathbb{E}(Z \mid \mathcal{F}_n)\mathbf{1}_A) \\ &= \lim_{n\to\infty} \mathbb{E}(\mathbb{E}(Z \mid \mathcal{F}_\infty)\mathbf{1}_A),\end{aligned}$$

where the last step relies on the fact that $\mathbf{1}_A$ is $\mathcal{F}_n$-measurable. □

We finally finish off the $p = 1$ case with the additional uniform integrability condition.

Theorem (Convergence in L^1)**.** Let $(X_n)_{n\geq 0}$ be a martingale. Then the following are equivalent:

(i) $(X_n)_{n\geq 0}$ is uniformly integrable.

(ii) $(X_n)_{n\geq 0}$ converges almost surely and in L^1.

(iii) There exists $Z \in L^1$ such that $X_n = \mathbb{E}(Z \mid \mathcal{F}_n)$ almost surely.

Moreover, $X_\infty = \mathbb{E}(Z \mid \mathcal{F}_\infty)$.

The proof is very similar to the L^p case.

Proof.

- (i) $\Rightarrow$ (ii): Let $(X_n)_{n\geq 0}$ be uniformly integrable. Then $(X_n)_{n\geq 0}$ is bounded in L^1. So the $(X_n)_{n\geq 0}$ converges to X_∞ almost surely. Then by measure theory, uniform integrability implies that in fact $X_n \to L^1$.
- (ii) $\Rightarrow$ (iii): Same as the L^p case.
- (iii) $\Rightarrow$ (i): For any $Z \in L^1$, the collection $\mathbb{E}(Z \mid \mathcal{G})$ ranging over all σ-subalgebras $\mathcal{G}$ is uniformly integrable. □

Thus, there is a bijection between uniformly integrable martingales and $L^1(\mathcal{F}_\infty)$.

We now revisit optional stopping for uniformly integrable martingales. Recall that in the statement of optional stopping, we needed our stopping times to be bounded. It turns out if we require our martingales to be uniformly integrable, then we can drop this requirement.

Theorem. If $(X_n)_{n\geq 0}$ is a uniformly integrable martingale, and S, T are arbitrary stopping times, then $\mathbb{E}(X_T \mid \mathcal{F}_S) = X_{S\wedge T}$. In particular $\mathbb{E}X_T = X_0$.

Note that we are now allowing arbitrary stopping times, so T may be infinite with non-zero probability. Hence we define

$$X_T = \sum_{n=0}^{\infty} X_n \mathbf{1}_{T=n} + X_\infty \mathbf{1}_{T=\infty}.$$

Proof. By optional stopping, for every n, we know that

$$\mathbb{E}(X_{T\wedge n} \mid \mathcal{F}_S) = X_{S\wedge T\wedge n}.$$

We want to be able to take the limit as $n \to \infty$. To do so, we need to show that things are uniformly integrable. First, we apply optional stopping to write $X_{T\wedge n}$ as

$$\begin{aligned} X_{T\wedge n} &= \mathbb{E}(X_n \mid \mathcal{F}_{T\wedge n}) \\ &= \mathbb{E}(\mathbb{E}(X_\infty \mid \mathcal{F}_n) \mid \mathcal{F}_{T\wedge n}) \\ &= \mathbb{E}(X_\infty \mid \mathcal{F}_{T\wedge n}). \end{aligned}$$

So we know $(X_n^T)_{n\geq 0}$ is uniformly integrable, and hence $X_{n\wedge T} \to X_T$ almost surely and in L^1.

To understand $\mathbb{E}(X_{T\wedge n} \mid \mathcal{F}_S)$, we note that

$$\|\mathbb{E}(X_{n\wedge T} - X_T \mid \mathcal{F}_S)\|_1 \leq \|X_{n\wedge T} - X_T\|_1 \to 0 \text{ as } n \to \infty.$$

So it follows that $\mathbb{E}(X_{n\wedge T} \mid \mathcal{F}_S) \to \mathbb{E}(X_T \mid \mathcal{F}_S)$ as $n \to \infty$. □

2.4 Applications of martingales

Having developed the theory, let us move on to some applications. Before we do that, we need the notion of a *backwards martingale*.

Definition (Backwards filtration)**.** A *backwards filtration* on a measurable space $(E, \mathcal{E})$ is a sequence of σ-algebras $\hat{\mathcal{F}}_n \subseteq \mathcal{E}$ such that $\hat{F}_{n+1} \subseteq \hat{F}_n$. We define

$$\hat{\mathcal{F}}_\infty = \bigcap_{n\geq 0} \hat{\mathcal{F}}_n.$$

Theorem. Let $Y \in L^1$, and let $\hat{\mathcal{F}}_n$ be a backwards filtration. Then

$$\mathbb{E}(Y \mid \hat{\mathcal{F}}_n) \to \mathbb{E}(Y \mid \hat{\mathcal{F}}_\infty)$$

almost surely and in L^1.

A process of this form is known as a *backwards martingale*.

Proof. We first show that $\mathbb{E}(Y \mid \hat{F}_n)$ converges. We then show that what it converges to is indeed $\mathbb{E}(Y \mid \hat{\mathcal{F}}_\infty)$.

We write

$$X_n = \mathbb{E}(Y \mid \hat{\mathcal{F}}_n).$$

Observe that for all $n \geq 0$, the process $(X_{n-k})_{0\leq k\leq n}$ is a martingale by the tower property, and so is $(-X_{n-k})_{0\leq k\leq n}$. Now notice that for all $a < b$, the number of upcrossings of $[a, b]$ by $(X_k)_{0\leq k\leq n}$ is equal to the number of upcrossings of $[-b, -a]$ by $(-X_{n-k})_{0\leq k\leq n}$.

Using the same arguments as for martingales, we conclude that $X_n \to X_\infty$ almost surely and in L^1 for some X_∞.

To see that $X_\infty = \mathbb{E}(Y \mid \hat{\mathcal{F}}_\infty)$, we notice that X_∞ is $\hat{F}_\infty$ measurable. So it is enough to prove that

$$\mathbb{E}X_\infty \mathbf{1}_A = \mathbb{E}(\mathbb{E}(Y \mid \hat{\mathcal{F}}_\infty)\mathbf{1}_A)$$

for all $A \in \hat{\mathcal{F}}_\infty$. Indeed, we have

$$\begin{aligned}
\mathbb{E}X_\infty \mathbf{1}_A &= \lim_{n\to\infty} \mathbb{E}X_n \mathbf{1}_A \\
&= \lim_{n\to\infty} \mathbb{E}(\mathbb{E}(Y \mid \hat{\mathcal{F}}_n)\mathbf{1}_A) \\
&= \lim_{n\to\infty} \mathbb{E}(Y \mid \mathbf{1}_A) \\
&= \mathbb{E}(Y \mid \mathbf{1}_A) \\
&= \mathbb{E}(\mathbb{E}(Y \mid \hat{\mathcal{F}}_n)\mathbf{1}_A).
\end{aligned}$$

□

Theorem (Kolmogorov 0-1 law)**.** Let $(X_n)_{n\geq 0}$ be independent random variables. Then, let

$$\hat{\mathcal{F}}_n = \sigma(X_{n+1}, X_{n+2}, \ldots).$$

Then the *tail σ-algebra* $\hat{\mathcal{F}}_\infty$ is trivial, i.e. $\mathbb{P}(A) \in \{0, 1\}$ for all $A \in \hat{\mathcal{F}}_\infty$.

Proof. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Then $\mathcal{F}_n$ and $\hat{\mathcal{F}}_n$ are independent. Then for all $A \in \hat{\mathcal{F}}_\infty$, we have

$$\mathbb{E}(\mathbf{1}_A \mid \mathcal{F}_n) = \mathbb{P}(A).$$

But the LHS is a martingale. So it converges almost surely and in L^1 to $\mathbb{E}(\mathbf{1}_A \mid \mathcal{F}_\infty)$. But $\mathbf{1}_A$ is $\mathcal{F}_\infty$-measurable, since $\hat{\mathcal{F}}_\infty \subseteq \mathcal{F}_\infty$. So this is just $\mathbf{1}_A$. So $\mathbf{1}_A = \mathbb{P}(A)$ almost surely, and we are done. □

Theorem (Strong law of large numbers)**.** Let $(X_n)_{n\geq 1}$ be iid random variables in L^1, with $\mathbb{E}X_1 = \mu$. Define

$$S_n = \sum_{i=1}^{n} X_i.$$

Then

$$\frac{S_n}{n} \to \mu \text{ as } n \to \infty$$

almost surely and in L^1.

Proof. We have

$$S_n = \mathbb{E}(S_n \mid S_n) = \sum_{i=1}^{n} \mathbb{E}(X_i \mid S_n) = n\mathbb{E}(X_1 \mid S_n).$$

So the problem is equivalent to showing that $\mathbb{E}(X_1 \mid S_n) \to \mu$ as $n \to \infty$. This seems like something we can tackle with our existing technology, except that the S_n do not form a filtration.

Thus, define a backwards filtration

$$\hat{\mathcal{F}}_n = \sigma(S_n, S_{n+1}, S_{n+2}, \ldots) = \sigma(S_n, X_{n+1}, X_{n+2}, \ldots) = \sigma(S_n, \tau_n),$$

where $\tau_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$. We now use the property of conditional expectation that we've never used so far, that adding independent information to a conditional expectation doesn't change the result. Since τ_n is independent of $\sigma(X_1, S_n)$, we know

$$\frac{S_n}{n} = \mathbb{E}(X_1 \mid S_n) = \mathbb{E}(X_1 \mid \hat{\mathcal{F}}_n).$$

Thus, by backwards martingale convergence, we know

$$\frac{S_n}{n} \to \mathbb{E}(X_1 \mid \hat{\mathcal{F}}_\infty).$$

But by the Kolmogorov 0-1 law, we know $\hat{\mathcal{F}}_\infty$ is trivial. So we know that $\mathbb{E}(X_1 \mid \hat{\mathcal{F}}_\infty)$ is almost constant, which has to be $\mathbb{E}(\mathbb{E}(X_1 \mid \hat{\mathcal{F}}_\infty)) = \mathbb{E}(X_1) = \mu$. □

Recall that if $(E, \mathcal{E}, \mu)$ is a measure space and $f \in m\mathcal{E}^+$, then

$$\nu(A) = \mu(f\mathbf{1}_A)$$

is a measure on $\mathcal{E}$. We say f is a density of ν with respect to μ.

We can ask an "inverse" question – given two different measures on $\mathcal{E}$, when is it the case that one is given by a density with respect to the other?

A first observation is that if $\nu(A) = \mu(f\mathbf{1}_A)$, then whenever $\mu(A) = 0$, we must have $\nu(A) = 0$. However, this is not sufficient. For example, let μ be a counting measure on $\mathbb{R}$, and ν the Lebesgue measure. Then our condition is satisfied. However, if ν is given by a density f with respect to ν, we must have

$$0 = \nu(\{x\}) = \mu(f\mathbf{1}_{\{x\}}) = f(x).$$

So $f \equiv 0$, but taking $f \equiv 0$ clearly doesn't give the Lebesgue measure.

The problem with this is that μ is not a σ-finite measure.

Theorem (Radon–Nikodym)**.** Let $(\Omega, \mathcal{F})$ be a measurable space, and $\mathbb{Q}$ and $\mathbb{P}$ be two probability measures on $(\Omega, \mathcal{F})$. Then the following are equivalent:

(i) $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$, i.e. for any $A \in \mathcal{F}$, if $\mathbb{P}(A) = 0$, then $\mathbb{Q}(A) = 0$.

(ii) For any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $A \in \mathcal{F}$, if $\mathbb{P}(A) \leq \delta$, then $\mathbb{Q}(A) \leq \varepsilon$.

(iii) There exists a random variable $X \geq 0$ such that

$$\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}(X \mathbf{1}_A).$$

In this case, X is called the *Radon–Nikodym derivative* of $\mathbb{Q}$ with respect to $\mathbb{P}$, and we write $X = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$.

Note that this theorem works for all finite measures by scaling, and thus for σ-finite measures by partitioning Ω into sets of finite measure.

Proof. We shall only treat the case where $\mathcal{F}$ is *countably generated*, i.e. $\mathcal{F} = \sigma(F_1, F_2, \ldots)$ for some sets F_i. For example, any second-countable topological space is countably generated.

- (iii) $\Rightarrow$ (i): Clear.
- (ii) $\Rightarrow$ (iii): Define the filtration

  $$\mathcal{F}_n = \sigma(F_1, F_2, \ldots, F_n).$$

  Since $\mathcal{F}_n$ is finite, we can write it as

  $$\mathcal{F}_n = \sigma(A_{n,1}, \ldots, A_{n,m_n}),$$

  where each $A_{n,i}$ is an *atom*, i.e. if $B \subsetneq A_{n,i}$ and $B \in \mathcal{F}_n$, then $B = \emptyset$. We define

  $$X_n = \sum_{n=1}^{m_n} \frac{\mathbb{Q}(A_{n,i})}{\mathbb{P}(A_{n,i})} \mathbf{1}_{A_{n,i}},$$

  where we skip over the terms where $\mathbb{P}(A_{n,i}) = 0$. Note that this is exactly designed so that for any $A \in \mathcal{F}_n$, we have

  $$\mathbb{E}_{\mathbb{P}}(X_n \mathbf{1}_A) = \mathbb{E}_{\mathbb{P}} \sum_{A_{n,i} \subseteq A} \frac{\mathbb{Q}(A_{n,i})}{\mathbb{P}(A_n, i)} \mathbf{1}_{A_{n,i}} = \mathbb{Q}(A).$$

  Thus, if $A \in \mathcal{F}_n \subseteq \mathcal{F}_{n+1}$, we have

  $$\mathbb{E} X_{n+1} \mathbf{1}_A = \mathbb{Q}(A) = \mathbb{E} X_n \mathbf{1}_A.$$

  So we know that

  $$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n.$$

  It is also immediate that $(X_n)_{n \geq 0}$ is adapted. So it is a martingale.

  We next show that $(X_n)_{n \geq 0}$ is uniformly integrable. By Markov's inequality, we have

  $$\mathbb{P}(X_n \geq \lambda) \leq \frac{\mathbb{E} X_n}{\lambda} = \frac{1}{\lambda} \leq \delta$$

for λ large enough. Then

$$\mathbb{E}(X_n \mathbf{1}_{X_n \geq \lambda}) = \mathbb{Q}(X_n \geq \lambda) \leq \varepsilon.$$

So we have shown uniform integrability, and so we know $X_n \to X$ almost surely and in L^1 for some X. Then for all $A \in \bigcup_{n\geq 0} \mathcal{F}_n$, we have

$$\mathbb{Q}(A) = \lim_{n\to\infty} \mathbb{E}X_n \mathbf{1}_A = \mathbb{E}X\mathbf{1}_A.$$

So $\mathbb{Q}(-)$ and $\mathbb{E}X\mathbf{1}_{(-)}$ agree on $\bigcup_{n\geq 0} \mathcal{F}_n$, which is a generating π-system for $\mathcal{F}$, so they must be the same.

– (i) $\Rightarrow$ (ii): Suppose not. Then there exists some $\varepsilon > 0$ and some $A_1, A_2, \ldots \in \mathcal{F}$ such that

$$\mathbb{Q}(A_n) \geq \varepsilon, \quad \mathbb{P}(A_n) \leq \frac{1}{2^n}.$$

Since $\sum_n \mathbb{P}(A_n)$ is finite, by Borel–Cantelli, we know

$$\mathbb{P} \limsup A_n = 0.$$

On the other hand, by, say, dominated convergence, we have

$$\begin{aligned}
\mathbb{Q} \limsup A_n &= \mathbb{Q}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) \\
&= \lim_{k\to\infty} \mathbb{Q}\left(\bigcap_{n=1}^{k} \bigcup_{m=n}^{\infty} A_m\right) \\
&\geq \lim_{k\to\infty} \mathbb{Q}\left(\bigcup_{m=k}^{\infty} A_k\right) \\
&\geq \varepsilon.
\end{aligned}$$

This is a contradiction. □

Finally, we end the part on discrete time processes by relating what we have done to Markov chains.

Let's first recall what Markov chains are. Let E be a countable space, and μ a measure on E. We write $\mu_x = \mu(\{x\})$, and then $\mu(f) = \mu \cdot f$.

Definition (Transition matrix)**.** A *transition matrix* is a matrix $P = (p_{xy})_{x,y\in E}$ such that each $p_x = (p_{x,y})_{y\in E}$ is a probability measure on E.

Definition (Markov chain)**.** An adapted process (X_n) is called a *Markov chain* if for any n and $A \in \mathcal{F}_n$ such that $\{x_n = x\} \supseteq A$, we have

$$\mathbb{P}(X_{n+1} = y \mid A) = p_{xy}.$$

Definition (Harmonic function)**.** A function $f : E \to \mathbb{R}$ is *harmonic* if $Pf = f$. In other words, for any x, we have

$$\sum_y p_{xy} f(y) = f(x).$$

We then observe that

Proposition. If F is harmonic and bounded, and $(X_n)_{n\geq 0}$ is Markov, then $(f(X_n))_{n\geq 0}$ is a martingale.

Example. Let $(X_n)_{n\geq 0}$ be iid $\mathbb{Z}$-valued random variables in L^1, and $\mathbb{E}[X_i] = 0$. Then

$$S_n = X_0 + \cdots + X_n$$

is a martingale and a Markov chain.

However, if Z is a $\mathbb{Z}$-valued random variable, consider the random variable $(ZS_n)_{n\geq 0}$ and $\mathcal{F}_n = \sigma(\mathcal{F}_n, Z)$. Then this is a martingale but not a Markov chain.

3 Continuous time stochastic processes

In the remainder of the course, we shall study continuous time processes. When doing so, we have to be rather careful, since our processes are index by an *uncountable* set, when measure theory tends to only like countable things. Ultimately, we would like to study Brownian motion, but we first develop some general theory of continuous time processes.

Definition (Continuous time stochastic process)**.** A *continuous time stochastic process* is a family of random variables $(X_t)_{t\geq 0}$ (or $(X_t)_{t\in[a,b]}$).

In the discrete case, if T is a random variable taking values in $\{0, 1, 2, \ldots\}$, then it makes sense to look at the new random variable X_T, since this is just

$$X_T = \sum_{n=0}^{\infty} X_n \mathbf{1}_{T=n}.$$

This is obviously measurable, since it is a limit of measurable functions.

However, this is not necessarily the case if we have continuous time, unless we assume some regularity conditions on our process. In some sense, we want X_t to depend "continuously" or at least "measurably" on t.

To make sense of X_T, It would be enough to require that the map

$$\varphi : (\omega, t) \mapsto X_t(\omega)$$

is measurable when we put the product σ-algebra on the domain. In this case, $X_T(\omega) = \varphi(\omega, T(\omega))$ is measurable. In this formulation, we see why we didn't have this problem with discrete time — the σ-algebra on $\mathbb{N}$ is just $\mathcal{P}(\mathbb{N})$, and so all sets are measurable. This is not true for $\mathcal{B}([0, \infty))$.

However, being able to talk about X_T is not the only thing we want. Often, the following definitions are useful:

Definition (Cadlag function)**.** We say a function $X : [0, \infty] \to \mathbb{R}$ is *cadlag* if for all t

$$\lim_{s\to t^+} x_s = x_t, \quad \lim_{s\to t^-} x_s \text{ exists.}$$

The name cadlag (or cádlág) comes from the French term *continue á droite, limite á gauche*, meaning "right-continuous with left limits".

Definition (Continuous/Cadlag stochastic process)**.** We say a stochastic process is *continuous* (resp. cadlag) if for any $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is continuous (resp. cadlag).

Notation. We write $C([0, \infty), \mathbb{R})$ for the space of all continuous functions $[0, \infty) \to \mathbb{R}$, and $D([0, \infty), \mathbb{R})$ the space of all cadlag functions.

We endow these spaces with a σ-algebra generated by the coordinate functions

$$(x_t)_{t\geq 0} \mapsto x_s.$$

Then a continuous (or cadlag) process is a random variable taking values in $C([0, \infty), \mathbb{R})$ (or $D([0, \infty), \mathbb{R})$).

Definition (Finite-dimensional distribution)**.** A *finite dimensional distribution* of $(X_t)_{t\geq 0}$ is a measure on $\mathbb{R}^n$ of the form

$$\mu_{t_1,\dots,t_n}(A) = \mathbb{P}((X_{t_1},\dots,X_{t_n}) \in A)$$

for all $A \in \mathcal{B}(\mathbb{R}^n)$, for some $0 \leq t_1 < t_2 < \dots < t_n$.

The important observation is that if we know all finite-dimensional distributions, then we know the law of X, since the cylinder sets form a π-system generating the σ-algebra.

If we know, a priori, that $(X_t)_{t\geq 0}$ is a continuous process, then for any dense set $I \subseteq [0,\infty)$, knowing $(X_t)_{t\geq 0}$ is the same as knowing $(X_t)_{t\in I}$. Conversely, if we are given some random variables $(X_t)_{t\in I}$, can we extend this to a continuous process $(X_t)_{t\geq 0}$? The answer is, of course, "not always", but it turns out we can if we assume some Hölder conditions.

Theorem (Kolmogorov's criterion)**.** Let $(\rho_t)_{t\in I}$ be random variables, where $I \subseteq [0,1]$ is dense. Assume that for some $p > 1$ and $\beta > \frac{1}{p}$, we have

$$\|\rho_t - \rho_s\|_p \leq C|t-s|^\beta \text{ for all } t,s \in I. \tag{$*$}$$

Then there exists a continuous process $(X_t)_{t\in I}$ such that for all $t \in I$,

$$X_t = \rho_t \text{ almost surely,}$$

and moreover for any $\alpha \in [0, \beta - \frac{1}{p})$, there exists a random variable $K_\alpha \in L^p$ such that

$$|X_s - X_t| \leq K_\alpha |s-t|^\alpha$$

for all $s,t \in [0,1]$.

Before we begin, we make the following definition:

Definition (Dyadic numbers)**.** We define

$$D_n = \left\{ s \in [0,1] : s = \frac{k}{2^n} \text{ for some } k \in \mathbb{Z} \right\}, \quad D = \bigcup_{n\geq 0} D_n.$$

Observe that $D \subseteq [0,1]$ is a dense subset. Topologically, this is just like any other dense subset. However, it is convenient to use D instead of an arbitrary subset when writing down formulas.

Proof. First note that we may assume $D \subseteq I$. Indeed, for $t \in D$, we can define ρ_t by taking the limit of ρ_s in L^p since L^p is complete. The equation $(*)$ is preserved by limits, so we may work on $I \cup D$ instead.

By assumption, $(\rho_t)_{t\in I}$ is Hölder in L^p. We claim that it is almost surely pointwise Hölder.

Claim. There exists a random variable $K_\alpha \in L^p$ such that

$$|\rho_s - \rho_t| \leq K_\alpha |s-t|^\alpha \text{ for all } s,t \in D.$$

Moreover, K_α is increasing in α.

Given the claim, we can simply set

$$X_t(\omega) = \begin{cases} \lim_{q\to t, q\in D} \rho_q(\omega) & K_\alpha < \infty \text{ for all } \alpha \in [0, \beta - \frac{1}{p}) \\ 0 & \text{otherwise} \end{cases}.$$

Then this is a continuous process, and satisfies the desired properties.

To construct such a K_α, observe that given any $s, t \in D$, we can pick $m \geq 0$ such that

$$2^{-(m+1)} < t - s \leq 2^{-m}.$$

Then we can pick $u = \frac{k}{2^{m+1}}$ such that $s < u < t$. Thus, we have

$$u - s < 2^{-m}, \quad t - u < 2^{-m}.$$

Therefore, by binary expansion, we can write

$$u - s = \sum_{i\geq m+1} \frac{x_i}{2^i}, \quad t - u = \sum_{i\geq m+1} \frac{y_i}{2^i},$$

for some $x_i, y_i \in \{0, 1\}$. Thus, writing

$$K_n = \sup_{t\in D_n} |S_{t+2^{-n}} - S_t|,$$

we can bound

$$|\rho_s - \rho_t| \leq 2 \sum_{n=m+1}^{\infty} K_n,$$

and thus

$$\frac{|\rho_s - \rho_t|}{|s-t|^\alpha} \leq 2 \sum_{n=m+1}^{\infty} 2^{(m+1)\alpha} K_n \leq 2 \sum_{n=m+1}^{\infty} 2^{(n+1)\alpha} K_n.$$

Thus, we can define

$$K_\alpha = 2 \sum_{n\geq 0} 2^{n\alpha} K_n.$$

We only have to check that this is in L^p, and this is not hard. We first get

$$\mathbb{E} K_n^p \leq \sum_{t\in D_n} \mathbb{E}|\rho_{t+2^{-n}} - \rho_t|^p \leq C^p 2^n \cdot 2^{-n\beta} = C 2^{n(1-p\beta)}.$$

Then we have

$$\|K_\alpha\|_p \leq 2 \sum_{n\geq 0} 2^{n\alpha} \|K_n\|_p \leq 2C \sum_{n\geq 0} 2^{n(\alpha + \frac{1}{p} - \beta)} < \infty. \qquad \square$$

We will later use this to construct Brownian motion. For now, we shall develop what we know about discrete time processes for continuous time ones. Fortunately, a lot of the proofs are either the same as the discrete time ones, or can be reduced to the discrete time version. So not much work has to be done!

Definition (Continuous time filtration)**.** A *continuous-time filtration* is a family of σ-algebras $(\mathcal{F}_t)_{t\geq 0}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ if $s \leq t$. Define $\mathcal{F}_\infty = \sigma(\mathcal{F}_t : t \geq 0)$.

Definition (Stopping time)**.** A random variable $t : \Omega \to [0, \infty]$ is a *stopping time* if $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

Proposition. Let $(X_t)_{t\geq 0}$ be a cadlag adapted process and S, T stopping times. Then

(i) $S \wedge T$ is a stopping time.

(ii) If $S \leq T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$.

(iii) $X_T \mathbf{1}_{T<\infty}$ is $\mathcal{F}_T$-measurable.

(iv) $(X_t^T)_{t\geq 0} = (X_{T\wedge t})_{t\geq 0}$ is adapted.

We only prove (iii). The first two are the same as the discrete case, and the proof of (iv) is similar to that of (iii).

To prove this, we need a quick lemma, whose proof is a simple exercise.

Lemma. A random variable Z is $\mathcal{F}_T$-measurable iff $Z\mathbf{1}_{\{T\leq t\}}$ is $\mathcal{F}_t$-measurable for all $t \geq 0$.

Proof of (iii) of proposition. We need to prove that $X_T \mathbf{1}_{\{T\leq t\}}$ is $\mathcal{F}_t$-measurable for all $t \geq 0$.

We write

$$X_T \mathbf{1}_{T\leq t} = X_T \mathbf{1}_{T<t} + X_t \mathbf{1}_{T=t}.$$

We know the second term is measurable. So it suffices to show that $X_T \mathbf{1}_{T<t}$ is $\mathcal{F}_t$-measurable.

Define $T_n = 2^{-n}\lceil 2^n T\rceil$. This is a stopping time, since we always have $T_n \geq T$. Since $(X_t)_{t\geq 0}$ is cadlag, we know

$$X_T \mathbf{1}_{T<t} = \lim_{n\to\infty} X_{T_n\wedge t}\mathbf{1}_{T<t}.$$

Now $T_n \wedge t$ can take only countably (and in fact only finitely) many values, so we can write

$$X_{T_n\wedge t} = \sum_{q\in D_n, q<t} X_q \mathbf{1}_{T_n=q} + X_t \mathbf{1}_{T<t<T_n},$$

and this is $\mathcal{F}_t$-measurable. So we are done. □

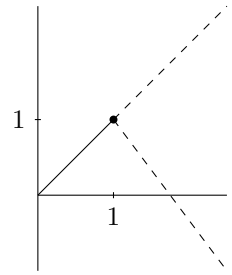
In the continuous case, stopping times are a bit more subtle. A natural source of stopping times is given by hitting times.

Definition (Hitting time)**.** Let $A \in \mathcal{B}(\mathbb{R})$. Then the *hitting time* of A is

$$T_A = \inf_{t\geq 0}\{X_t \leq A\}.$$

This is not always a stopping time. For example, consider the process X_t such that with probability $\frac{1}{2}$, it is given by $X_t = t$, and with probability $\frac{1}{2}$, it is given by

$$X_t = \begin{cases} t & t \leq 1 \\ 2 - t & t > 1 \end{cases}.$$



Take $A = (1, \infty)$. Then $T_A = 1$ in the first case, and $T_A = \infty$ in the second case. But $\{T_a \leq 1\} \notin \mathcal{F}_1$, as at time 1, we don't know if we are going up or down.

The problem is that A is not closed.

Proposition. Let $A \subseteq \mathbb{R}$ be a closed set and $(X_t)_{t\geq 0}$ be continuous. Then T_A is a stopping time.

Proof. Observe that $d(X_q, A)$ is a continuous function in q. So we have

$$\{T_A \leq t\} = \left\{\inf_{q\in\mathbb{Q}, q<t} d(X_q, A) = 0\right\}. \qquad \square$$

Motivated by our previous non-example of a hitting time, we define

Definition (Right-continuous filtration)**.** Given a continuous filtration $(\mathcal{F}_t)_{t\geq 0}$, we define

$$\mathcal{F}_t^+ = \bigcap_{s>t} \mathcal{F}_s \supseteq \mathcal{F}_t.$$

We say $(\mathcal{F}_t)_{t\geq 0}$ is *right continuous* if $\mathcal{F}_t = \mathcal{F}_t^+$.

Often, we want to modify our events by things of measure zero. While this doesn't really affect anything, it could potentially get us out of $\mathcal{F}_t$. It does no harm to enlarge all $\mathcal{F}_t$ to include events of measure zero.

Definition (Usual conditions)**.** Let $\mathcal{N} = \{A \in \mathcal{F}_\infty : \mathbb{P}(A) \in \{0, 1\}\}$. We say that $(\mathcal{F}_t)_{t\geq 0}$ satisfies the *usual conditions* if it is right continuous and $\mathcal{N} \subseteq \mathcal{F}_0$.

Proposition. Let $(X_t)_{t\geq 0}$ be an adapted process (to $(\mathcal{F}_t)_{t\geq 0}$) that is cadlag, and let A be an open set. Then T_A is a stopping time with respect to $\mathcal{F}_t^+$.

Proof. Since $(X_t)_{t\geq 0}$ is cadlag and A is open. Then

$$\{T_A < t\} = \bigcup_{q<t, q\in\mathbb{Q}} \{X_q \in A\} \in \mathcal{F}_t.$$

Then

$$\{T_A \leq t\} = \bigcap_{n\geq 0} \left\{T_A < t + \frac{1}{n}\right\} \in \mathcal{F}_t^+. \qquad \square$$

Definition (Coninuous time martingale)**.** An adapted process $(X_t)_{t\geq 0}$ is called a *martingale* iff

$$\mathbb{E}(X_t \mid \mathcal{F}_s) = X_s$$

for all $t \geq s$, and similarly for super-martingales and sub-martingales.

Note that if $t_1 \leq t_2 \leq \cdots$, then

$$\tilde{X}_n = X_{t_n}$$

is a discrete time martingale. Similarly, if $t_1 \geq t_2 \geq \cdots$, and

$$\hat{X}_n = X_{t_n}$$

defines a discrete time backwards martingale. Using this observation, we can now prove what we already know in the discrete case.

Theorem (Optional stopping theorem)**.** Let $(X_t)_{t\geq 0}$ be an adapted cadlag process in L^1. Then the following are equivalent:

(i) For any bounded stopping time T and any stopping time S, we have $X_T \in L^1$ and
$$\mathbb{E}(X_T \mid \mathcal{F}_S) = X_{T\wedge S}.$$

(ii) For any stopping time T, $(X_t^T)_{t\geq 0} = (X_{T\wedge t})_{t\geq 0}$ is a martingale.

(iii) For any bounded stopping time T, $X_T \in L^1$ and $\mathbb{E}X_T = \mathbb{E}X_0$.

Proof. We show that (i) $\Rightarrow$ (ii), and the rest follows from the discrete case similarly.

Since T is bounded, assume $T \leq t$, and we may wlog assume $t \in \mathbb{N}$. Let

$$T_n = 2^{-n}\lceil 2^n T\rceil, \quad S_n = 2^{-n}\lceil 2^n S\rceil.$$

We have $T_n \searrow T$ as $n \to \infty$, and so $X_{T_n} \to X_T$ as $n \to \infty$.

Since $T_n \leq t + 1$, by restricting our sequence to D_n, discrete time optional stopping implies

$$\mathbb{E}(X_{t+1} \mid \mathcal{F}_{T_n}) = X_{T_n}.$$

In particular, X_{T_n} is uniformly integrable. So it converges in L^1. This implies $X_T \in L^1$.

To show that $\mathbb{E}(X_t \mid \mathcal{F}_S) = X_{T\wedge S}$, we need to show that for any $A \in \mathcal{F}_S$, we have

$$\mathbb{E}X_t\mathbf{1}_A = \mathbb{E}X_{S\wedge T}\mathbf{1}_A.$$

Since $\mathcal{F}_S \subseteq \mathcal{F}_{S_n}$, we already know that

$$\mathbb{E}X_{T_n}\mathbf{1}_A = \lim_{n\to\infty} \mathbb{E}X_{S_n\wedge T_n}\mathbf{1}_A$$

by discrete time optional stopping, since $\mathbb{E}(X_{T_n} \mid \mathcal{F}_{S_n}) = X_{T_n\wedge S_n}$. So taking the limit $n \to \infty$ gives the desired result. □

Theorem. Let $(X_t)_{t\geq 0}$ be a super-martingale bounded in L^1. Then it converges almost surely as $t \to \infty$ to a random variable $X_\infty \in L^1$.

Proof. Define $U_s[a, b, (x_t)_{t\geq 0}]$ be the number of upcrossings of $[a, b]$ by $(x_t)_{t\geq 0}$ up to time s, and

$$U_\infty[a, b, (x_t)_{t\geq 0}] = \lim_{s\to\infty} U_s[a, b, (x_t)_{t\geq 0}].$$

Then for all $s \geq 0$, we have

$$U_s[a, b, (x_t)_{t\geq 0}] = \lim_{n\to\infty} U_s[a, b, (x_t^{(n)})_{t\in D_n}].$$

By monotone convergence and Doob's upcrossing lemma, we have

$$\mathbb{E}U_s[a, b, (X_t)_{t\geq 0}] = \lim_{n\to\infty} \mathbb{E}U_s[a, b, (X_t)_{t\in D_n}] \leq \frac{\mathbb{E}(X_s - a)^-}{b-1} \leq \frac{\mathbb{E}|X_s| + a}{b-a}.$$

We are then done by taking the supremum over s. Then finish the argument as in the discrete case.

This shows we have pointwise convergence in $\mathbb{R} \cup \{\pm\infty\}$, and by Fatou's lemma, we know that

$$\mathbb{E}|X_\infty| = \mathbb{E}\liminf_{t_n\to\infty} |X_{t_n}| \leq \liminf_{t_n\to\infty} \mathbb{E}|X_{t_n}| < \infty.$$

So X_∞ is finite almost surely. □

We shall now state without proof some results we already know for the discrete case. The proofs are straightforward generalizations of the discrete version.

Lemma (Maximal inequality)**.** Let $(X_t)_{t\geq 0}$ be a cadlag martingale or a non-negative sub-martingale. Then for all $t \geq 0$, $\lambda \geq 0$, we have

$$\lambda\mathbb{P}(X_t^* \geq \lambda) \leq \mathbb{E}|X_t|.$$

Lemma (Doob's L^p inequality)**.** Let $(X_t)_{t\geq 0}$ be as above. Then

$$\|X_t^*\|_p \leq \frac{p}{p-1}\|X_t\|_p.$$

Definition (Version)**.** We say a process $(Y_t)_{t\geq 0}$ is a *version* of $(X_t)_{t\geq 0}$ if for all t, $\mathbb{P}(Y_t = X_t) = 1$.

Note that this not the same as saying $\mathbb{P}(\forall_t : Y_t = X_t) = 1$.

Example. Take $X_t \equiv 0$ for all t and take U be a uniform random variable on $[0, 1]$. Define

$$Y_t = \begin{cases} 1 & t = U \\ 0 & \text{otherwise} \end{cases}.$$

Then for all t, we have $X_t = Y_t$ almost surely. So (Y_t) is a version of (X_t). However, X_t is continuous but Y_t is not.

Theorem (Regularization of martingales)**.** Let $(X_t)_{t\geq 0}$ be a martingale with respect to $(\mathcal{F}_t)$, and suppose $\mathcal{F}_t$ satisfies the usual conditions. Then there exists a version $(\tilde{X}_t)$ of (X_t) which is cadlag.

Proof. For all $M > 0$, define

$$\Omega_0^M = \left\{\sup_{q\in D\cap[0,M]} |X_q| < \infty\right\} \cap \bigcap_{a<b\in\mathbb{Q}} \left\{U_M[a, b, (X_t)_{t\in D\cap[0,M]}] < \infty\right\}$$

Then we see that $\mathbb{P}(\Omega_0^M) = 1$ by Doob's upcrossing lemma. Now define

$$\tilde{X}_t = \lim_{s \geq t, s \to t, s \in D} X_s \mathbf{1}_{\Omega_0^t}.$$

Then this is $\mathcal{F}_t$ measurable because $\mathcal{F}_t$ satisfies the usual conditions.

Take a sequence $t_n \searrow t$. Then (X_{t_n}) is a backwards martingale. So it converges almost surely in L^1 to $\tilde{X}_t$. But we can write

$$X_t = \mathbb{E}(X_{t_n} \mid \mathcal{F}_t).$$

Since $X_{t_n} \to \tilde{X}_t$ in L^1, and $\tilde{X}_t$ is $\mathcal{F}_t$-measurable, we know $X_t = \tilde{X}_t$ almost surely.

The fact that it is cadlag is an exercise. □

Theorem (L^p convergence of martingales)**.** Let $(X_t)_{t \geq 0}$ be a cadlag martingale. Then the following are equivalent:

(i) $(X_t)_{t \geq 0}$ is bounded in L^p.

(ii) $(X_t)_{t \geq 0}$ converges almost surely and in L^p.

(iii) There exists $Z \in L^p$ such that $X_t = \mathbb{E}(Z \mid \mathcal{F}_t)$ almost surely.

Theorem (L^1 convergence of martingales)**.** Let $(X_t)_{t \geq 0}$ be a cadlag martingale. Then the folloiwng are equivalent:

(i) $(X_t)_{t \geq 0}$ is uniformly integrable.

(ii) $(X_t)_{t \geq 0}$ converges almost surely and in L^1 to X_∞.

(iii) There exists $Z \in L^1$ such that $\mathbb{E}(Z \mid \mathcal{F}_t) = X_t$ almost surely.

Theorem (Optional stopping theorem)**.** Let $(X_t)_{t \geq 0}$ be a uniformly integrable martingale, and let S, T b e any stopping times. Then

$$\mathbb{E}(X_T \mid \mathcal{F}_s) = X_{S \wedge T}.$$

4 Weak convergence of measures

Often, we may want to consider random variables defined on different spaces. Since we cannot directly compare them, a sensible approach would be to use them to push our measure forward to $\mathbb{R}$, and compare them on $\mathbb{R}$.

Definition (Law)**.** Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. The *law* of X is the probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by

$$\mu(A) = \mathbb{P}(X^{-1}(A)).$$

Example. For $x \in \mathbb{R}$, we have the *Dirac δ measure*

$$\delta_x(A) = \mathbf{1}_{\{x \in A\}}.$$

This is the law of a random variable that constantly takes the value x.

Now if we have a sequence $x_n \to x$, then we would like to say $\delta_{x_n} \to \delta_x$. In what sense is this true? Suppose f is continuous. Then

$$\int f \mathrm{d}\delta_{x_n} = f(x_n) \to f(x) = \int f \mathrm{d}\delta_x.$$

So we do have some sort of convergence if we pair it with a continuous function.

Definition (Weak convergence)**.** Let $(\mu_n)_{n\geq 0}$, μ be probability measures on a metric space (M, d) with the Borel measure. We say that $\mu_n \Rightarrow \mu$, or μ_n *converges weakly* to μ if

$$\mu_n(f) \to \mu(f)$$

for all f bounded and continuous.

If $(X_n)_{n\geq 0}$ are random variables, then we say (X_n) converges *in distribution* if μ_{X_n} converges weakly.

Note that in general, weak convergence does not say anything about how measures of subsets behave.

Example. If $x_n \to x$, then $\delta_{x_n} \to \delta_x$ weakly. However, if $x_n \neq x$ for all n, then $\delta_{x_n}(\{x\}) = 0$ but $\delta_x(\{x\}) = 1$. So

$$\delta_{x_n}(\{x\}) \not\to \delta_n(\{x\}).$$

Example. Pick $X = [0, 1]$. Let $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{\frac{k}{n}}$. Then

$$\mu_n(f) = \frac{1}{n}\sum_{k=1}^n f\left(\frac{k}{n}\right).$$

So μ_n converges to the Lebesgue measure.

Proposition. Let $(\mu_n)_{n\geq 0}$ be as above. Then, the following are equivalent:

(i) $(\mu_n)_{n\geq 0}$ converges weakly to μ.

(ii) For all open G, we have

$$\liminf_{n\to\infty} \mu_n(G) \geq \mu(G).$$

(iii) For all closed A, we have

$$\limsup_{n\to\infty} \mu_n(A) \le \mu(A).$$

(iv) For all A such that $\mu(\partial A) = 0$, we have

$$\lim_{n\to\infty} \mu_n(A) = \mu(A)$$

(v) (when $M = \mathbb{R}$) $F_{\mu_n}(x) \to F_\mu(x)$ for all x at which F_μ is continuous, where F_μ is the *distribution function* of μ, defined by $F_\mu(x) = \mu_n((-\infty, t])$.

Proof.

- (i) $\Rightarrow$ (ii): The idea is to approximate the open set by continuous functions. We know A^c is closed. So we can define

  $$f_N(x) = 1 \wedge (N \cdot \operatorname{dist}(x, A^c)).$$

  This has the property that for all $N > 0$, we have

  $$f_N \le \mathbf{1}_A,$$

  and moreover $f_N \nearrow \mathbf{1}_A$ as $N \to \infty$. Now by definition of weak convergence,

  $$\liminf_{n\to\infty} \mu(A) \ge \liminf_{n\to\infty} \mu_n(f_N) = \mu(F_N) \to \mu(A) \text{ as } N \to \infty.$$

- (ii) $\Leftrightarrow$ (iii): Take complements.

- (iii) and (ii) $\Rightarrow$ (iv): Take A such that $\mu(\partial A) = 0$. Then

  $$\mu(A) = \mu(\mathring{A}) = \mu(\bar{A}).$$

  So we know that

  $$\liminf_{n\to\infty} \mu_n(A) \ge \liminf_{n\to\infty} \mu_n(\mathring{A}) \ge \mu(\mathring{A}) = \mu(A).$$

  Similarly, we find that

  $$\mu(A) \ge \limsup_{n\to\infty} \mu_n(A).$$

  So we are done.

- (iv) $\Rightarrow$ (i): We have

  $$\begin{aligned}\mu(f) &= \int_M f(x)\, \mathrm{d}\mu(x)\\ &= \int_M \int_0^\infty \mathbf{1}_{f(x)\ge t}\, \mathrm{d}t\, \mathrm{d}\mu(x)\\ &= \int_0^\infty \mu(\{f \ge t\})\, \mathrm{d}t.\end{aligned}$$

  Since f is continuous, $\partial\{f \le t\} \subseteq \{f = t\}$. Now there can be only countably many t's such that $\mu(\{f = t\}) > 0$. So replacing μ by $\lim_{n\to\infty} \mu_n$ only changes the integrand at countably many places, hence doesn't affect the integral. So we conclude using bounded convergence theorem.

– (iv) ⇒ (v): Assume t is a continuity point of F_μ. Then we have

$$\mu(\partial(-\infty, t]) = \mu(\{t\}) = F_\mu(t) - F_\mu(t_-) = 0.$$

So $\mu_n(\partial_n(-\infty, t]) \to \mu((-\infty, t])$, and we are done.

– (v) ⇒ (ii): If $A = (a, b)$, then

$$\mu_n(A) \geq F_{\mu_n}(b') - F_{\mu_n}(a')$$

for any $a \leq a' \leq b' \leq b$ with a', b' continuity points of F_μ. So we know that

$$\liminf_{n\to\infty} \mu_n(A) \geq F_\mu(b') - F_\mu(a') = \mu(a', b').$$

By taking supremum over all such a', b', we find that

$$\liminf_{n\to\infty} \mu_n(A) \geq \mu(A). \qquad \square$$

Definition (Tight probability measures)**.** A sequence of probability measures $(\mu_n)_{n\geq 0}$ on a metric space (M, e) is *tight* if for all $\varepsilon > 0$, there exists compact $K \subseteq M$ such that

$$\sup_n \mu_n(M \setminus K) \leq \varepsilon.$$

Note that this is always satisfied for compact metric spaces.

Theorem (Prokhorov's theorem)**.** If $(\mu_n)_{n\geq 0}$ is a sequence of tight probability measures, then there is a subsequence $(\mu_{n_k})_{k\geq 0}$ and a measure μ such that $\mu_{n_k} \Rightarrow \mu$.

To see how this can fail without the tightness assumption, suppose we define measures μ_n on $\mathbb{R}$ by

$$\mu_n(A) = \tilde{\mu}(A \cap [n, n+1]),$$

where $\tilde{\mu}$ is the Lebesgue measure. Then for any bounded set S, we have $\lim_{n\to\infty} \mu_n(S) = 0$. Thus, if the weak limit existed, it must be everywhere zero, but this does not give a probability measure.

We shall prove this only in the case $M = \mathbb{R}$. It is not difficult to construct a candidate of what the weak limit should be. Simply use Bolzano–Weierstrass to pick a subsequence of the measures such that the distribution functions converge on the rationals. Then the limit would essentially be what we want. We then apply tightness to show that this is a genuine distribution.

Proof. Take $\mathbb{Q} \subseteq \mathbb{R}$, which is dense and countable. Let $x_1, x_2, \ldots$ be an enumeration of $\mathbb{Q}$. Define $F_n = F_{\mu_n}$. By Bolzano–Weierstrass, and some fiddling around with sequences, we can find some F_{n_k} such that

$$F_{n_k}(x_i) \to y_i \equiv F(x_i)$$

as $k \to \infty$, for each fixed x_i.

Since F is non-decreasing on $\mathbb{Q}$, it has left and right limits everywhere. We extend F to $\mathbb{R}$ by taking right limits. This implies F is cadlag.

Take x a continuity point of F. Then for each $\varepsilon > 0$, there exists $s < x < t$ rational such that

$$|F(s) - F(t)| < \frac{\varepsilon}{2}.$$

Take n large enough such that $|F_n(s) - F(s)| < \frac{\varepsilon}{4}$, and same for t. Then by monotonicity of F and F_n, we have

$$|F_n(x) - F(x)| \le |F(s) - F(t)| + |F_n(s) - F(s)| + |F_n(t) - F(t)| \le \varepsilon.$$

It remains to show that $F(x) \to 1$ as $x \to \infty$ and $F(x) \to 0$ as $x \to -\infty$. By tightness, for all $\varepsilon > 0$, there exists $N > 0$ such that

$$\mu_n((-\infty, N]) \le \varepsilon, \quad \mu_n((N, \infty) \le \varepsilon.$$

This then implies what we want. □

We shall end the chapter with an alternative characterization of weak convergence, using characteristic functions.

Definition (Characteristic function)**.** Let X be a random variable taking values in $\mathbb{R}^d$. The *characteristic function* of X is the function $\mathbb{R}^d \to \mathbb{C}$ defined by

$$\varphi_X(t) = \mathbb{E}e^{i\langle t, x\rangle} = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} \, \mathrm{d}\mu_X(x).$$

Note that φ_X is continuous by bounded convergence, and $\varphi_X(0) = 1$.

Proposition. If $\varphi_X = \varphi_Y$, then $\mu_X = \mu_Y$.

Theorem (Lévy's convergence theroem)**.** Let $(X_n)_{n\ge 0}$, X be random variables taking values in $\mathbb{R}^d$. Then the following are equivalent:

(i) $\mu_{X_n} \Rightarrow \mu_X$ as $n \to \infty$.

(ii) $\varphi_{X_n} \to \varphi_X$ pointwise.

We will in fact prove a stronger theorem.

Theorem (Lévy)**.** Let $(X_n)_{n\ge 0}$ be as above, and let $\varphi_{X_n}(t) \to \psi(t)$ for all t. Suppose ψ is continuous at 0 and $\psi(0) = 1$. Then there exists a random variable X such that $\varphi_X = \psi$ and $\mu_{X_n} \Rightarrow \mu_X$ as $n \to \infty$.

We will only prove the case $d = 1$. We first need the following lemma:

Lemma. Let X be a real random variable. Then for all $\lambda > 0$,

$$\mu_X(|x| \ge \lambda) \le c\lambda \int_0^{1/\lambda} (1 - \operatorname{Re}\varphi_X(t)) \, \mathrm{d}t,$$

where $C = (1 - \sin 1)^{-1}$.

Proof. For $M \ge 1$, we have

$$\int_0^M (1 - \cos t) \, \mathrm{d}t = M - \sin M \ge M(1 - \sin 1).$$

By setting $M = \frac{|x|}{\lambda}$, we have

$$\mathbf{1}_{|X|\ge\lambda} \le C\frac{\lambda}{|X|} \int_0^{|X|/\lambda} (1 - \cos t) \, \mathrm{d}t.$$

By a change of variables with $t \mapsto Xt$, we have

$$\mathbf{1}_{|X|\geq\lambda} \leq c\lambda \int_0^1 (1 - \cos Xt)\, \mathrm{d}t.$$

Apply μ_X, and use the fact that $\operatorname{Re}\varphi_X(t) = \mathbb{E}\cos(Xt)$. □

We can now prove Lévy's theorem.

Proof of theorem. It is clear that weak convergence implies convergence in characteristic functions.

Now observe that if $\mu_n \Rightarrow \mu$ iff from every subsequence $(n_k)_{k\geq 0}$, we can choose a further subsequence (n_{k_ℓ}) such that $\mu_{n_{k_\ell}} \Rightarrow \mu$ as $\ell \to \infty$. Indeed, $\Rightarrow$ is clear, and suppose $\mu_n \not\Rightarrow \mu$ but satisfies the subsequence property. Then we can choose a bounded and continuous function f such that

$$\mu_n(f) \not\to \mu(f).$$

Then there is a subsequence $(n_k)_{k\geq 0}$ such that $|\mu_{n_k}(f) - \mu(f)| > \varepsilon$. Then there is no further subsequence that converges.

Thus, to show $\Leftarrow$, we need to prove the existence of subsequential limits (uniqueness follows from convergence of characteristic functions). It is enough to prove tightness of the whole sequence.

By the mean value theorem, we can choose λ so large that

$$c\lambda \int_0^{1/\lambda} (1 - \operatorname{Re}\psi(t))\, \mathrm{d}t < \frac{\varepsilon}{2}.$$

By bounded convergence, we can choose λ so large that

$$c\lambda \int_0^{1/\lambda} (1 - \operatorname{Re}\varphi_{X_n}(t))\, \mathrm{d}t \leq \varepsilon$$

for all n. Thus, by our previous lemma, we know $(\mu_{X_n})_{n\geq 0}$ is tight. So we are done. □

5 Brownian motion

Finally, we can begins studying Brownian motion. Brownian motion was first observed by the botanist Robert Brown in 1827, when he looked at the random movement of pollen grains in water. In 1905, Albert Einstein provided the first mathematical description of this behaviour. In 1923, Norbert Wiener provided the first rigorous construction of Brownian motion.

5.1 Basic properties of Brownian motion

Definition (Brownian motion)**.** A continuous process $(B_t)_{t\geq 0}$ taking values in $\mathbb{R}^d$ is called a *Brownian motion* in $\mathbb{R}^d$ started at $x \in \mathbb{R}^d$ if

(i) $B_0 = x$ almost surely.

(ii) For all $s < t$, the *increment* $B_t - B_s \sim N(0, (t-s)I)$.

(iii) Increments are independent. More precisely, for all $t_1 < t_2 < \cdots < t_k$, the random variables

$$B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}}$$

are independent.

If $B_0 = 0$, then we call it a *standard Brownian motion*.

We always assume our Brownian motion is standard.

Theorem (Wiener's theorem)**.** There exists a Brownian motion on some probability space.

Proof. We first prove existence on $[0, 1]$ and in $d = 1$. We wish to apply Kolmogorov's criterion.

Recall that D_n are the dyadic numbers. Let $(Z_d)_{d\in D}$ be iid $N(0, 1)$ random variables on some probability space. We will define a process on D_n inductively on n with the required properties. We wlog assume $x = 0$.

In step 0, we put

$$B_0 = 0, \quad B_1 = Z_1.$$

Assume that we have already constructed $(B_d)_{d\in D_{n-1}}$ satisfying the properties. Take $d \in D_n \setminus D_{n-1}$, and set

$$d^{\pm} = d \pm 2^{-n}.$$

These are the two consecutive numbers in D_{n-1} such that $d_- < d < d_+$. Define

$$B_d = \frac{B_{d_+} + B_{d_-}}{2} + \frac{1}{2^{(n+1)/2}} Z_d.$$

The condition (i) is trivially satisfied. We now have to check the other two conditions.

Consider

$$\begin{aligned} B_{d_+} - B_d &= \frac{B_{d_+} - B_{d_-}}{2} - \frac{1}{2^{(n+1)/2}} Z_d \\ B_d - B_{d_-} &= \underbrace{\frac{B_{d_+} - B_{d_-}}{2}}_{N} + \underbrace{\frac{1}{2^{(n+1)/2}} Z_d}_{N'}. \end{aligned}$$

Notice that N and N' are normal with variance $\mathrm{var}(N') = \mathrm{var}(N) = \frac{1}{2^{n+1}}$. In particular, we have

$$\mathrm{cov}(N - N', N + N') = \mathrm{var}(N) - \mathrm{var}(N') = 0.$$

So $B_{d_+} - B_d$ and $B_d - B_{d_-}$ are independent.

Now note that the vector of increments of $(B_d)_{d\in D_n}$ between consecutive numbers in D_n is Gaussian, since after dotting with any vector, we obtain a linear combination of independent Gaussians. Thus, to prove independence, it suffice to prove that pairwise correlation vanishes.

We already proved this for the case of increments between B_d and $B_{d_\pm}$, and this is the only case that is tricky, since they both involve the same Z_d. The other cases are straightforward, and are left as an exercise for the reader.

Inductively, we can construct $(B_d)_{d\in D}$, satisfying (i), (ii) and (iii). Note that for all $s, t \in D$, we have

$$\mathbb{E}|B_t - B_s|^p = |t - s|^{p/2}\mathbb{E}|N|^p$$

for $N \sim N(0, 1)$. Since $\mathbb{E}|N|^p < \infty$ for all p, by Kolmogorov's criterion, we can extend $(B_d)_{d\in D}$ to $(B_t)_{t\in[0,1]}$. In fact, this is α-Hölder continuous for all $\alpha < \frac{1}{2}$.

Since this is a continuous process and satisfies the desired properties on a dense set, it remains to show that the properties are preserved by taking continuous limits.

Take $0 \le t_1 < t_2 < \cdots < t_m \le 1$, and $0 \le t_1^n < t_2^n < \cdots < t_m^n \le 1$ such that $t_i^n \in D_n$ and $t_i^n \to t_i$ as $n \to \infty$ and $i = 1, \dots m$.

We now apply Lévy's convergence theorem. Recall that if X is a random variable in $\mathbb{R}^d$ and $X \sim N(0, \Sigma)$, then

$$\varphi_X(u) = \exp\left(-\frac{1}{2}u^T\Sigma u\right).$$

Since $(B_t)_{t\in[0,1]}$ is continuous, we have

$$\begin{aligned}\varphi_{(B_{t_2^n} - B_{t_1^n}, \dots, B_{t_m^n} - B_{t_{m-1}^n})}(u) &= \exp\left(-\frac{1}{2}u^T\Sigma u\right)\\ &= \exp\left(-\frac{1}{2}\sum_{i=1}^{m-1}(t_{i+1}^n - t_i^n)u_i^2\right).\end{aligned}$$

We know this converges, as $n \to \infty$, to $\exp\left(-\frac{1}{2}\sum_{i=1}^{m-1}(t_{i+1} - t_i)u_i^2\right)$.

By Lévy's convergence theorem, the law of $(B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \dots, B_{t_n} - B_{t_{m-1}})$ is Gaussian with the right covariance. This implies that (ii) and (iii) hold on $[0, 1]$.

To extend the time to $[0, \infty)$, we define independent Brownian motions $(B_t^i)_{t\in[0,1], i\in\mathbb{N}}$ and define

$$B_t = \sum_{i=0}^{\lfloor t\rfloor - 1} B_1^i + B_{t-\lfloor t\rfloor}^{\lfloor t\rfloor}$$

To extend to $\mathbb{R}^d$, take the product of d many independent one-dimensional Brownian motions. □

Lemma. Brownian motion is a Gaussian process, i.e. for any $0 \leq t_1 < t_2 < \cdots < t_m \leq 1$, the vector $(B_{t_1}, B_{t_2}, \ldots, B_{t_n})$ is Gaussian with covariance

$$\operatorname{cov}(B_{t_1}, B_{t_2}) = t_1 \wedge t_2.$$

Proof. We know $(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_m} - B_{t_{m-1}})$ is Gaussian. Thus, the sequence $(B_{t_1}, \ldots, B_{t_m})$ is the image under a linear isomorphism, so it is Gaussian. To compute covariance, for $s \leq t$, we have

$$\operatorname{cov}(B_s, B_t) = \mathbb{E}B_s B_t = \mathbb{E}B_s B_T - \mathbb{E}B_s^2 + \mathbb{E}B_s^2 = \mathbb{E}B_s(B_t - B_s) + \mathbb{E}B_s^2 = s. \ \square$$

Proposition (Invariance properties)**.** Let $(B_t)_{t\geq 0}$ be a standard Brownian motion in $\mathbb{R}^d$.

(i) If U is an orthogonal matrix, then $(UB_t)_{t\geq 0}$ is a standard Brownian motion.

(ii) *Brownian scaling*: If $a > 0$, then $(a^{-1/2}B_{at})_{t\geq 0}$ is a standard Brownian motion. This is known as a *random fractal property*.

(iii) (*Simple*) *Markov property*: For all $s \geq 0$, the sequence $(B_{t+s} - B_s)_{t\geq 0}$ is a standard Brownian motion, independent of $(\mathcal{F}_s^B)$.

(iv) *Time inversion*: Define a process

$$X_t = \begin{cases} 0 & t = 0 \\ tB_{1/t} & t > 0 \end{cases}.$$

Then $(X_t)_{t\geq 0}$ is a standard Brownian motion.

Proof. Only (iv) requires proof. It is enough to prove that X_t is continuous and has the right finite-dimensional distributions. We haves

$$(X_{t_1}, \ldots, X_{t_m}) = (t_1 B_{1/t_1}, \ldots, t_m B_{1/t_m}).$$

The right-hand side is the image of $(B_{1/t_1}, \ldots, B_{1/t_m})$ under a linear isomorphism. So it is Gaussian. If $s \leq t$, then the covariance is

$$\operatorname{cov}(sB_s, tB_t) = st \operatorname{cov}(B_{1/s}, B_{1/t}) = st\left(\frac{1}{s} \wedge \frac{1}{t}\right) = s = s \wedge t.$$

Continuity is obvious for $t > 0$. To prove continuity at 0, we already proved that $(X_q)_{q>0, q\in\mathbb{Q}}$ has the same law (as a process) as Brownian motion. By continuity of X_t for positive t, we have

$$\mathbb{P}\left(\lim_{q\in\mathbb{Q}_+, q\to 0} X_q = 0\right) = \mathbb{P}\left(\lim_{q\in\mathbb{Q}_+, q\to 0} B_q = 0\right) = \mathbf{1}_B$$

by continuity of B. $\square$

Using the natural filtration, we have

Theorem. For all $s \geq t$, the process $(B_{t+s} - B_s)_{t\geq 0}$ is independent of $\mathcal{F}_s^+$.

Proof. Take a sequence $s_n \to s$ such that $s_n > s$ for all n. By continuity,

$$B_{t+s} - B_s = \lim_{n\to\infty} B_{t+s_n} - B_{s_n}$$

almost surely. Now each of $B_{t+s_n} - B_{s_n}$ is independent of $\mathcal{F}_s^+$, and hence so is the limit. □

Theorem (Blumenthal's 0-1 law)**.** The σ-algebra $\mathcal{F}_0^+$ is trivial, i.e. if $A \in \mathcal{F}_0^+$, then $\mathbb{P}(A) \in \{0, 1\}$.

Proof. Apply our previous theorem. Take $A \in \mathcal{F}_0^+$. Then $A \in \sigma(B_s : s \geq 0)$. So A is independent of itself. □

Proposition.

(i) If $d = 1$, then

$$\begin{aligned} 1 &= \mathbb{P}(\inf\{t \geq 0 : B_t > 0\} = 0) \\ &= \mathbb{P}(\inf\{t \geq 0 : B_t < 0\} = 0) \\ &= \mathbb{P}(\inf\{t > 0 : B_t = 0\} = 0) \end{aligned}$$

(ii) For any $d \geq 1$, we have

$$\lim_{t\to\infty} \frac{B_t}{t} = 0$$

almost surely.

(iii) If we define

$$S_t = \sup_{0\leq s\leq t} B_t, \quad I_t = \inf_{0\leq s\leq t} B_t,$$

then $S_\infty = \infty$ and $I_\infty = -\infty$ almost surely.

(iv) If A is open an $\mathbb{R}^d$, then the cone of A is $C_A = \{tx : x \in A, t > 0\}$. Then $\inf\{t \geq 0 : B_t \in C_A\} = 0$ almost surely.

Thus, Brownian motion is pretty chaotic.

Proof.

(i) It suffices to prove the first equality. Note that the event $\{\inf\{t \geq 0 : B_k > 0\} = 0\}$ is trivial. Moreover, for any finite t, the probability that $B_t > 0$ is $\frac{1}{2}$. Then take a sequence t_n such that $t_n \to 0$, and apply Fatou to conclude that the probability is positive.

(ii) Follows from the previous one since $tB_{1/t}$ is a Brownian motion.

(iii) By scale invariance, because $S_\infty = aS_\infty$ for all $a > 0$.

(iv) Same as (i). □

Theorem (Strong Markov property)**.** Let $(B_t)_{t\geq 0}$ be a standard Brownian motion in $\mathbb{R}^d$, and let T be an almost-surely finite stopping time with respect to $(\mathcal{F}_t^+)_{t\geq 0}$. Then

$$\tilde{B}_t = B_{T+t} - B_T$$

is a standard Brownian motion with respect to $(\mathcal{F}_{T+t}^+)_{t\geq 0}$ that is independent of $\mathcal{F}_T^+$.

Proof. Let $T_n = 2^{-n}\lceil 2^n T\rceil$. We first prove the statement for T_n. We let

$$B_t^{(k)} = B_{t+k/2^n} - B_{k/2^n}$$

This is then a standard Browninan motion independent of $\mathcal{F}^+_{k/2^n}$ by the simple Markov property. Let

$$B_*(t) = B_{t+T_n} - B_{T_n}.$$

Let $\mathcal{A}$ be the σ-algebra on $\mathcal{C} = C([0,\infty), \mathbb{R}^d)$, and $A \in \mathcal{A}$. Let $E \in \mathcal{F}^+_{T_n}$. The claim that B_* is a standard Brownian motion independent of E can be concisely captured in the equality

$$\mathbb{P}(\{B_* \in A\} \cap E) = \mathbb{P}(\{B \in A\})\mathbb{P}(E). \tag{$\dagger$}$$

Taking $E = \Omega$ tells us B_* and B have the same law, and then taking general E tells us B_* is independent of $\mathcal{F}^+_{T_n}$.

It is a straightforward computation to prove ($\dagger$). Indeed, we have

$$\mathbb{P}(\{B_* \in A\} \cap E) = \sum_{k=0}^\infty \mathbb{P}\left(\{B^{(k)} \in A\} \cap E \cap \left\{T_n = \frac{k}{2^n}\right\}\right)$$

Since $E \in \mathcal{F}^+_{T_n}$, we know $E \cap \{T_n = k/2^n\} \in \mathcal{F}^+_{k/2^n}$. So by the simple Markov property, this is equal to

$$= \sum_{k=0}^\infty \mathbb{P}(\{B^{(k)} \in A\})\mathbb{P}\left(E \cap \left\{T_n = \frac{k}{2^n}\right\}\right).$$

But we know B_k is a standard Brownian motion. So this is equal to

$$\begin{aligned}
&= \sum_{b=0}^\infty \mathbb{P}(\{B \in A\})\mathbb{P}\left(E \cap \left\{T_n = \frac{k}{2^n}\right\}\right)\\
&= \mathbb{P}(\{B \in A\})\mathbb{P}(E).
\end{aligned}$$

So we are done.

Now as $n \to \infty$, the increments of B_* converge almost surely to the increments of $\tilde{B}$, since B is continuous and $T_n \searrow T$ almost surely. But B_* all have the same distribution, and almost sure convergence implies convergence in distribution. So $\tilde{B}$ is a standard Brownian motion. Being independent of $\mathcal{F}^+_T$ is clear. □

We know that we can reset our process any time we like, and we also know that we have a bunch of invariance properties. We can combine these to prove some nice results.

Theorem (Reflection principle)**.** Let $(B_t)_{T\geq 0}$ and T be as above. Then the *reflected process* $(\tilde{B}_t)_{t\geq 0}$ defined by

$$\tilde{B}_t = B_t \mathbf{1}_{t<T} + (2B_T - B_t)\mathbf{1}_{t\geq T}$$

is a standard Brownian motion.

Of course, the fact that we are reflecting is not important. We can apply any operation that preserves the law. This theorem is "obvious", but we can be a bit more careful in writing down a proof.

Proof. By the strong Markov property, we know

$$B_t^T = B_{T+t} - B_T$$

and $-B_t^T$ are standard Brownian motions independent of $\mathcal{F}_T^+$. This implies that the pairs of random variables

$$P_1 = ((B_t)_{0\le t\le T}, (B_t)_{t\ge 0}^T), \quad P_2 = ((B_t)_{0\le t\le T}, (-B_t)_{t\ge 0}^T)$$

taking values in $\mathcal{C}\times\mathcal{C}$ have the same law on $\mathcal{C}\times\mathcal{C}$ with the product σ-algebra.

Define the concatenation map $\psi_T(X,Y) : \mathcal{C}\times\mathcal{C}\to\mathcal{C}$ by

$$\psi_T(X,Y) = X_t\mathbf{1}_{t<T} + (X_T + Y_{t-T})\mathbf{1}_{t\ge T}.$$

Assuming $Y_0 = 0$, the resulting process is continuous.

Notice that ψ_T is a measurable map, which we can prove by approximations of T by discrete stopping times. We then conclude that $\psi_T(P_1)$ has the same law as $\psi_T(P_2)$. □

Corollary. Let $(B_t)_{T\ge 0}$ be a standard Brownian motion in $d = 1$. Let $b > 0$ and $a \le b$. Let

$$S_t = \sup_{0\le s\le t} B_t.$$

Then

$$\mathbb{P}(S_t \ge b, B_t \le a) = \mathbb{P}(B_t \ge 2b - a).$$

Proof. Consider the stopping time T given by the first hitting time of b. Since $S_\infty = \infty$, we know T is finite almost surely. Let $(\tilde{B}_t)_{t\ge 0}$ be the reflected process. Then

$$\{S_t \ge b, B_t \le a\} = \{\tilde{B}_t \ge 2b - a\}.$$ □

Corollary. The law of S_t is equal to the law of $|B_t|$.

Proof. Apply the previous process with $b = a$ to get

$$\begin{aligned}\mathbb{P}(S_t \ge a) &= \mathbb{P}(S_t \ge a, B_t < a) + \mathbb{P}(S_t \ge a, B_t \ge a)\\ &= \mathbb{P}(B_t \ge a) + \mathbb{P}(B_t \ge a)\\ &= \mathbb{P}(B_t \le a) + \mathbb{P}(B_t \ge a)\\ &= \mathbb{P}(|B_t| \ge a).\end{aligned}$$ □

Proposition. Let $d = 1$ and $(B_t)_{t\ge 0}$ be a standard Brownian motion. Then the following processes are $(\mathcal{F}_t^+)_{t\ge 0}$ martingales:

(i) $(B_t)_{t\ge 0}$

(ii) $(B_t^2 - t)_{t\ge 0}$

(iii) $\left(\exp\left(uB_t - \frac{u^2 t}{2}\right)\right)_{t\ge 0}$ for $u \in \mathbb{R}$.

Proof.

(i) Using the fact that $B_t - B_s$ is independent of $\mathcal{F}_s^+$, we know

$$\mathbb{E}(B_t - B_s \mid \mathcal{F}_s^+) = \mathbb{E}(B_t - B_s) = 0.$$

(ii) We have

$$\mathbb{E}(B_t^2 - t \mid \mathcal{F}_s^+) = \mathbb{E}((B_t - B_s)^2 \mid \mathcal{F}_s) - \mathbb{E}(B_s^2 \mid \mathcal{F}_s^+) + 2\mathbb{E}(B_t B_s \mid \mathcal{F}_s^+) - t$$

We know $B_t - B_s$ is independent of $\mathcal{F}_s^+$, and so the first term is equal to $\mathrm{var}(B_t - B_s) = (t - s)$, and we can simply to get

$$\begin{aligned} &= (t - s) - B_s^2 + 2B_s^2 - t \\ &= B_s^2 - s. \end{aligned}$$

(iii) Similar. □

5.2 Harmonic functions and Brownian motion

Recall that a Markov chain plus a harmonic function gave us a martingale. We shall derive similar results here.

Definition (Domain)**.** A *domain* is an open connected set $D \subseteq \mathbb{R}^d$.

Definition (Harmonic function)**.** A function $u : D \to \mathbb{R}$ is called *harmonic* if

$$\Delta f = \sum_{i=1}^{d} \frac{\partial^2 f}{\partial x_i^2} = 0.$$

There is also an alternative characterization of harmonic functions that involves integrals instead of derivatives.

Lemma. Let $u : D \to \mathbb{R}$ be measurable and locally bounded. Then the following are equivalent:

(i) u is twice continuously differentiable and $\Delta u = 0$.

(ii) For any $x \in D$ and $r > 0$ such that $B(x, r) \subseteq D$, we have

$$u(x) = \frac{1}{\mathrm{Vol}(B(x, r))} \int_{B(x,r)} u(y)\, \mathrm{d}y$$

(iii) For any $x \in D$ and $r > 0$ such that $B(x, r) \subseteq D$, we have

$$u(x) = \frac{1}{\mathrm{Area}(\partial B(x, r))} \int_{\partial B(x,r)} u(y)\, \mathrm{d}y.$$

The latter two properties are known as the *mean value property*.

Proof. IA Vector Calculus. □

Theorem. Let $(B_t)_{t \geq 0}$ be a standard Brownian motion in $\mathbb{R}^d$, and $u : \mathbb{R}^d \to \mathbb{R}$ be harmonic such that

$$\mathbb{E}|u(x + B_t)| < \infty$$

for any $x \in \mathbb{R}^d$ and $t \geq 0$. Then the process $(u(B_t))_{t \geq 0}$ is a martingale with respect to $(\mathcal{F}_t^+)_{t \geq 0}$.

To prove this, we need to prove a side lemma:

Lemma. If X and Y are independent random variables in $\mathbb{R}^d$, and X is $\mathcal{G}$-measurable. If $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is such that $f(X, Y)$ is integrable, then

$$\mathbb{E}(f(X, Y) \mid \mathcal{G}) = \mathbb{E}f(z, Y)|_{z=X}.$$

Proof. Use Fubini and the fact that $\mu_{(X,Y)} = \mu_X \otimes \mu_Y$. □

Observe that if μ is a probability measure in $\mathbb{R}^d$ such that the density of μ with respect to the Lebesgue measure depends only on $|x|$, then if u is harmonic, the mean value property implies

$$u(x) = \int_{\mathbb{R}^d} u(x + y)\, \mathrm{d}\mu(y).$$

Proof of theorem. Let $t \geq s$. Then

$$\begin{aligned}
\mathbb{E}(u(B_t) \mid \mathcal{F}_s^+) &= \mathbb{E}(u(B_s + (B_t - B_s)) \mid \mathcal{F}_s^+) \\
&= \mathbb{E}(u(z + B_t - B_s))|_{Z=B_s} \\
&= u(z)|_{z=B_s} \\
&= u(B_s).
\end{aligned}$$

□

In fact, the following more general result is true:

Theorem. Let $f : \mathbb{R}^d \to \mathbb{R}$ be twice continuously differentiable with bounded derivatives. Then, the processes $(X_t)_{t\geq 0}$ defined by

$$X_t = f(B_t) - \frac{1}{2}\int_0^t \Delta f(B_s)\, \mathrm{d}s$$

is a martingale with respect to $(\mathcal{F}_t^+)_{t\geq 0}$.

We shall not prove this, but we can justify this as follows: suppose we have a sequence of independent random variables $\{X_1, X_2, \ldots\}$, with

$$\mathbb{P}(X_i = \pm 1) = \frac{1}{2}.$$

Let $S_n = X_1 + \cdots + X_n$. Then

$$\mathbb{E}(f(S_{n+1}) \mid S_1, \ldots, S_n) - f(S_n) = \frac{1}{2}(f(S_n - 1) + f(S_n + 1) - 2f(s_n)) \equiv \frac{1}{2}\tilde{\Delta} f(S_n),$$

and we see that this is the discretized second derivative. So

$$f(S_n) - \frac{1}{2}\sum_{i=0}^{n-1} \tilde{\Delta} f(S_i)$$

is a martingale.

Now the mean value property of a harmonic function u says if we draw a sphere B centered at x, then $u(x)$ is the average value of u on B. More generally, if we have a surface S containing x, is it true that $u(x)$ is the average value of u on S in some sense?

Remarkably, the answer is yes, and the precise result is given by Brownian motion. Let $(X_t)_{t\geq 0}$ be a Brownian motion started at x, and let T be the first hitting time of S. Then, under certain technical conditions, we have

$$u(x) = \mathbb{E}_x u(X_T).$$

In fact, given some function φ defined on the boundary of D, we can set

$$u(x) = \mathbb{E}_x \varphi(X_T),$$

and this gives us the (unique) solution to Laplace's equation with the boundary condition given by φ.

It is in fact not hard to show that the resulting u is harmonic in D, since it is almost immediate by construction that $u(x)$ is the average of u on a small sphere around x. The hard part is to show that u is in fact continuous at the boundary, so that it is a genuine solution to the boundary value problem. This is where the technical condition comes in.

First, we quickly establish that solutions to Laplace's equation are unique.

Definition (Maximum principle)**.** Let $u : \bar{D} \to \mathbb{R}$ be continuous and harmonic. Then

(i) If u attains its maximum inside D, then u is constant.

(ii) If D is bounded, then the maximum of u in $\bar{D}$ is attained at ∂D.

Thus, harmonic functions do not have interior maxima unless it is constant.

Proof. Follows from the mean value property of harmonic functions. □

Corollary. If u and u' solve $\Delta u = \Delta u' = 0$, and u and u' agree on ∂D, then $u = u'$.

Proof. $u - u'$ is also harmonic, and so attains the maximum at the boundary, where it is 0. Similarly, the minimum is attained at the boundary. □

The technical condition we impose on D is the following:

Definition (Poincaré cone condition)**.** We say a domain D satisfies the *Poincaré cone condition* if for any $x \in \partial D$, there is an open cone C based at X such that

$$C \cap D \cap B(x, \delta) = \emptyset$$

for some $\delta \geq 0$.

Example. If $D = \mathbb{R}^2 \setminus (\{0\} \times \mathbb{R}_{\geq 0})$, then D does not satisfy the Poincaré cone condition.

And the technical lemma is as follows:

Lemma. Let C be an open cone in $\mathbb{R}^d$ based at 0. Then there exists $0 \leq a < 1$ such that if $|x| \leq \frac{1}{2^k}$, then

$$\mathbb{P}_x(T_{\partial B(0,1)} < T_C) \leq a^k.$$

Proof. Pick

$$a = \sup_{|x| \le \frac{1}{2}} \mathbb{P}_x(T_{\partial B(0,1)} < T_C) < 1.$$

We then apply the strong Markov property, and the fact that Brownian motion is scale invariant. We reason as follows — if we start with $|x| \le \frac{1}{2^k}$, we may or may not hit $\partial B(2^{-k+1})$ before hitting C. If we don't, then we are happy. If we are not, then we have reached $\partial B(2^{-k+1})$. This happens with probability at most a. Now that we are at $\partial B(2^{-k+1})$, the probability of hitting $\partial B(2^{-k+2})$ before hitting the cone is at most a again. If we hit $\partial B(2^{-k+3})$, we again have a probability of $\le a$ of hitting $\partial B(2^{-k+4})$, and keep going on. Then by induction, we find that the probability of hitting $\partial B(0,1)$ before hitting the cone is $\le a^k$. □

The ultimate theorem is then

Theorem. Let D be a bounded domain satisfying the Poincaré cone condition, and let $\varphi : \partial D \to \mathbb{R}$ be continuous. Let

$$T_{\partial D} = \inf\{t \ge 0 : B_t \in \partial D\}.$$

This is a *bounded* stopping time. Then the function $u : \bar{D} \to \mathbb{R}$ defined by

$$u(x) = \mathbb{E}_x(\varphi(B_{T_{\partial D}})),$$

where $\mathbb{E}_x$ is the expectation if we start at x, is the unique continuous function such that $u(x) = \varphi(x)$ for $x \in \partial D$, and $\Delta u = 0$ for $x \in D$.

Proof. Let $\tau = T_{\partial B(x,\delta)}$ for δ small. Then by the strong Markov property and the tower law, we have

$$u(x) = \mathbb{E}_x(u(x_\tau)),$$

and x_τ is uniformly distributed over $\partial B(x, \delta)$. So we know u is harmonic in the interior of D, and in particular is continuous in the interior. It is also clear that $u|_{\partial D} = \varphi$. So it remains to show that u is continuous up to $\bar{D}$.

So let $x \in \partial D$. Since φ is continuous, for every $\varepsilon > 0$, there is $\delta > 0$ such that if $y \in \partial D$ and $|y - x| < \delta$, then $|\varphi(y) - \varphi(x)| \le \varepsilon$.

Take $z \in \bar{D}$ such that $|z - x| \le \frac{\delta}{2}$. Suppose we start our Brownian motion at z. If we hit the boundary before we leave the ball, then we are in good shape. If not, then we are sad. But if the second case has small probability, then since φ is bounded, we can still be fine.

Pick a cone C as in the definition of the Poincaré cone condition, and assume we picked δ small enough that $C \cap B(x, \delta) \cap D = \emptyset$. Then we have

$$\begin{aligned}
|u(z) - \varphi(x)| &= |\mathbb{E}_z(\varphi(B_{T_{\partial D}})) - \varphi(x)| \\
&\le \mathbb{E}_z|\varphi(B_{T_{\partial D}} - \varphi(x))| \\
&\le \varepsilon \mathbb{P}_z(T_{B(x,\delta)} > T_{\partial D}) + 2 \sup \|\varphi\| \mathbb{P}_z(T_{\partial D} > T_{\partial B(x,\delta)}) \\
&\le \varepsilon + 2\|\varphi\|_\infty \mathbb{P}_z(T_{B(x,\delta)} \le T_C),
\end{aligned}$$

and we know the second term $\to 0$ as $z \to x$. □

5.3 Transience and recurrence

Theorem. Let $(B_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$.

- If $d = 1$, then $(B_t)_{t\geq 0}$ is *point recurrent*, i.e. for each $x, z \in \mathbb{R}$, the set $\{t \geq 0 : B_t = z\}$ is unbounded $\mathbb{P}_x$-almost surely.
- If $d = 2$, then $(B_t)_{t\geq 0}$ is *neighbourhood recurrent*, i.e. for each $x \in \mathbb{R}^2$ and $U \subseteq \mathbb{R}^2$ open, the set $\{t \geq 0 : B_t \in U\}$ is unbounded $\mathbb{P}_x$-almost surely. However, the process does not visit points, i.e. for all $x, z \in \mathbb{R}^d$, we have
$$\mathbb{P}_X(B_t = z \text{ for some } t > 0) = 0.$$
- If $d \geq 3$, then $(B_t)_{t\geq 0}$ is *transient*, i.e. $|B_t| \to \infty$ as $t \to \infty$ $\mathbb{P}_x$-almost surely.

Proof.

- This is trivial, since $\inf_{t\geq 0} B_t = -\infty$ and $\sup_{t\geq 0} B_t = \infty$ almost surely, and $(B_t)_{t\geq 0}$ is continuous.
- It is enough to prove for $x = 0$. Let $0 < \varepsilon < R < \infty$ and $\varphi \in C_b^2(\mathbb{R}^2)$ such that $\varphi(x) = \log|x|$ for $\varepsilon \leq |x| \leq R$. It is an easy exercise to check that this is harmonic inside the annulus. By the theorem we didn't prove, we know
$$M_t = \varphi(B_t) - \frac{1}{2}\int_0^t \Delta\varphi(B_s)\,\mathrm{d}s$$
is a martingale. For $\lambda \geq 0$, let $S_\lambda = \inf\{t \geq 0 : |B_t| = \lambda\}$. If $\varepsilon \leq |x| \leq R$, then $H = S_\varepsilon \wedge S_R$ is $\mathbb{P}_X$-almost surely finite. Then M_H is a bounded martingale. By optional stopping, we have
$$\mathbb{E}_x(\log|B_H|) = \log|x|.$$
But the LHS is
$$\log\varepsilon\mathbb{P}(S_\varepsilon < S_R) + \log R\mathbb{P}(S_R < S_\varepsilon).$$
So we find that
$$\mathbb{P}_x(S_\varepsilon < S_R) = \frac{\log R - \log|x|}{\log R - \log\varepsilon}. \qquad (*)$$
Note that if we let $R \to \infty$, then $S_R \to \infty$ almost surely. Using $(*)$, this implies $\mathbb{P}_X(S_\varepsilon < \infty) = 1$, and this does not depend on x. So we are done.

  To prove that $(B_t)_{t\geq 0}$ does not visit points, let $\varepsilon \to 0$ in $(*)$ and then $R \to \infty$ for $x \neq z = 0$.
- It is enough to consider the case $d = 3$. As before, let $\varphi \in C_b^2(\mathbb{R}^3)$ be such that
$$\varphi(x) = \frac{1}{|x|}$$
for $\varepsilon \leq x \leq 2$. Then $\Delta\varphi(x) = 0$ for $\varepsilon \leq x \leq R$. As before, we get
$$\mathbb{P}_x(S_\varepsilon < S_R) = \frac{|x|^{-1} - |R|^{-1}}{\varepsilon^{-1} - R^{-1}}.$$

As $R \to \infty$, we have

$$\mathbb{P}_x(S_\varepsilon < \infty) = \frac{\varepsilon}{x}.$$

Now let

$$A_n = \{B_t \geq n \text{ for all } t \geq B_{T_{n^3}}\}.$$

Then

$$\mathbb{P}_0(A_n^c) = \frac{1}{n^2}.$$

So by Borel–Cantelli, we know only finitely of A_n^c occur almost surely. So infinitely many of the A_n hold. This guarantees our process $\to \infty$. □

5.4 Donsker's invariance principle

To end our discussion on Brownian motion, we provide an alternative construction of Brownian motion, given by Donsker's invariance principle. Suppose we run any simple random walk on $\mathbb{Z}^d$. We can think of this as a very coarse approximation of a Brownian motion. As we zoom out, the step sizes in the simple random walk look smaller and smaller, and if we zoom out sufficiently much, then we might expect that the result looks like Brownian motion, and indeed it converges to Brownian motion in the limit.

Theorem (Donsker's invariance principle)**.** Let $(X_n)_{n\geq 0}$ be iid random variables with mean 0 and variance 1, and set $S_n = X_1 + \cdots + X_n$. Define

$$S_t = (1 - \{t\})s_{\lfloor t \rfloor} + \{t\}S_{\lfloor t \rfloor + 1}.$$

where $\{t\} = t - \lfloor t \rfloor$.

Define

$$(S_t^{[N]})_{t\geq 0} = (N^{-1/2} S_{t\cdot N})_{t\in[0,1]}.$$

As $(S_t^{[N]})_{t\in[0,1]}$ converges in distribution to the law of standard Brownian motion on $[0, 1]$.

The reader might wonder why we didn't construct our Brownian motion this way instead of using Wiener's theorem. The answer is that our proof of Donsker's invariance principle relies on the existence of a Brownian motion! The relevance is the following theorem:

Theorem (Skorokhod embedding theorem)**.** Let μ be a probability measure on $\mathbb{R}$ with mean 0 and variance σ^2. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\geq 0}$ on which there is a standard Brownian motion $(B_t)_{t\geq 0}$ and a sequence of stopping times $(T_n)_{n\geq 0}$ such that, setting $S_n = B_{T_n}$,

(i) T_n is a random walk with steps of mean σ^2

(ii) S_n is a random walk with step distribution μ.

So in some sense, Brownian motion contains all random walks with finite variance.

The only stopping times we know about are the hitting times of some value. However, if we take T_n to be the hitting time of some fixed value, then B_{T_n} would be a pretty poor attempt at constructing a random walk. Thus, we may try to come up with the following strategy — construct a probability space

with a Brownian motion $(B_t)_{t\geq 0}$, and an independent iid sequence $(X_n)_{n\in\mathbb{N}}$ of random variables with distribution μ. We then take T_n to be the first hitting time of $X_1 + \cdots + X_n$. Then setting $S_n = X_{T_n}$, property (ii) is by definition satisfied. However, (i) will not be satisfied in general. In fact, for any $y \neq 0$, the expected first hitting time of y is infinite! The problem is that if, say, $y > 0$, and we accidentally strayed off to the negative side, then it could take a long time to return.

The solution is to "split" μ into two parts, and construct two random variables $(X, Y) \in [0, \infty)^2$, such that if T is T is the first hitting time of $(-X, Y)$, then B_T has law μ.

Since we are interested in the stopping times T_{-x} and T_y, the following computation will come in handy:

Lemma. Let $x, y > 0$. Then

$$\mathbb{P}_0(T_{-x} < T_y) = \frac{y}{x+y}, \quad \mathbb{E}_0 T_{-x} \wedge T_y = xy.$$

Proof sketch. Use optional stopping with $(B_t^2 - t)_{t\geq 0}$. □

Proof of Skorokhod embedding theorem. Define Borel measures $\mu^{\pm}$ on $[0, \infty)$ by

$$\mu^{\pm}(A) = \mu(\pm A).$$

Note that these are not probability measures, but we can define a probability measure ν on $[0, \infty)^2$ given by

$$\mathrm{d}\nu(x, y) = C(x + y)\, \mathrm{d}\mu^-(x)\, \mathrm{d}\mu^+(y)$$

for some normalizing constant C (this is possible since μ is integrable). This $(x+y)$ is the same $(x+y)$ appearing in the denominator of $\mathbb{P}_0(T_{-x} < T_y) = \frac{y}{x+y}$. Then we claim that any (X, Y) with this distribution will do the job.

We first figure out the value of C. Note that since μ has mean 0, we have

$$C \int_0^\infty x\, \mathrm{d}\mu^-(x) = C \int_0^\infty y\, \mathrm{d}\mu^+(y).$$

Thus, we have

$$\begin{aligned}
1 &= \int C(x + y)\, \mathrm{d}\mu^-(x)\, \mathrm{d}\mu^+(y)\\
&= C \int x\, \mathrm{d}\mu^-(x) \int \mathrm{d}\mu^+(y) + C \int y\, \mathrm{d}\mu^+(y) \int \mathrm{d}\mu^-(x)\\
&= C \int x\, \mathrm{d}\mu^-(x) \left(\int \mathrm{d}\mu^+(y) + \int \mathrm{d}\mu^-(x) \right)\\
&= C \int x\, \mathrm{d}\mu^-(x) = C \int y\, \mathrm{d}\mu^+(y).
\end{aligned}$$

We now set up our notation. Take a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a standard Brownian motion $(B_t)_{t\geq 0}$ and a sequence $((X_n, Y_n))_{n\geq 0}$ iid with distribution ν and independent of $(B_t)_{t\geq 0}$.

Define

$$\mathcal{F}_0 = \sigma((X_n, Y_n), n = 1, 2, \ldots), \quad \mathcal{F}_t = \sigma(\mathcal{F}_0, \mathcal{F}_t^B).$$

Define a sequence of stopping times

$$T_0 = 0, T_{n+1} = \inf\{t \geq T_n : B_t - B_{T_n} \in \{-X_{n+1}, Y_{n+1}\}\}.$$

By the strong Markov property, it suffices to prove that things work in the case $n = 1$. So for convenience, let $T = T_1, X = X_1, Y = Y_1$.

To simplify notation, let $\tau : C([0,1], \mathbb{R}) \times [0,\infty)^2 \to [0,\infty)$ be given by

$$\tau(\omega, x, y) = \inf\{t \geq 0 : \omega(t) \in \{-x, y\}\}.$$

Then we have

$$T = \tau((B_t)_{t\geq 0}, X, Y).$$

To check that this works, i.e. (ii) holds, if $A \subseteq [0,\infty)$, then

$$\mathbb{P}(B_T \in A) = \int_{[0,\infty)^2} \int_{C([0,\infty),\mathbb{R})} \mathbf{1}_{\tau(\omega,x,y)\in A}\, \mathrm{d}\mu_B(\omega)\, \mathrm{d}\nu(x,y).$$

Using the first part of the previous computation, this is given by

$$\int_{[0,\infty)^2} \frac{x}{x+y} \mathbf{1}_{y\in A}\, C(x+y)\, \mathrm{d}\mu^-(x)\, \mathrm{d}\mu^+(y) = \mu^+(A).$$

We can prove a similar result if $A \subseteq (-\infty, 0)$. So B_T has the right law.

To see that T is also well-behaved, we compute

$$\begin{aligned}
\mathbb{E}T &= \int_{[0,\infty)^2} \int_{\mathbb{C}([0,1],\mathbb{R})} \tau(\omega, x, y)\, \mathrm{d}\mu_B(\omega)\, \mathrm{d}\nu(x,y)\\
&= \int_{[0,\infty)^2} xy\, \mathrm{d}\nu(x,y)\\
&= C\int_{[0,\infty)^2} (x^2 y + yx^2)\, \mathrm{d}\mu^-(x)\, \mathrm{d}\mu^+(y)\\
&= \int_{[0,\infty)} x^2\, \mathrm{d}\mu^-(x) + \int_{[0,\infty)} y^2\, \mathrm{d}\mu^+(y)\\
&= \sigma^2.
\end{aligned}$$

□

The idea of the proof of Donsker's invariance principle is that in the limit of large N, the T_n are roughly regularly spaced, by the law of large numbers, so this allows us to reverse the above and use the random walk to approximate the Brownian motion.

Proof of Donsker's invariance principle. Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. Then by Brownian scaling,

$$(B_t^{(N)})_{t\geq 0} = (N^{1/2} B_{t/N})_{t\geq 0}$$

is a standard Brownian motion.

For every $N > 0$, we let $(T_n^{(N)})_{n\geq 0}$ be a sequence of stopping times as in the embedding theorem for $B^{(N)}$. We then set

$$S_n^{(N)} = B^{(N)}_{T_n^{(N)}}.$$

For t not an integer, define $S_t^{(N)}$ by linear interpolation. Observe that

$$((T_n^{(N)})_{n\geq 0}, S_t^{(N)}) \sim ((T_n^{(1)})_{n\geq 0}, S_t^{(1)}).$$

We define

$$\tilde{S}_t^{(N)} = N^{-1/2} S_{tN}^{(N)}, \quad \tilde{T}_n^{(N)} = \frac{T_n^{(N)}}{N}.$$

Note that if $t = \frac{n}{N}$, then

$$\tilde{S}_{n/N}^{(N)} = N^{-1/2} S_n^{(N)} = N^{-1/2} B_{T_n^{(N)}}^{(N)} = B_{T_n^{(N)}/N} = B_{\tilde{T}_n^N}. \tag{$*$}$$

Note that $(\tilde{S}_t^{(N)})_{t\geq 0} \sim (S_t^{(N)})_{t\geq 0}$. We will prove that we have convergence in probability, i.e. for any $\delta > 0$,

$$\mathbb{P}\left(\sup_{0\leq t<1} |\tilde{S}_t^{(N)} - B_t| > \delta\right) = \mathbb{P}(\|\tilde{S}^{(N)} - B\|_\infty > \delta) \to 0 \text{ as } N \to \infty.$$

We already know that $\tilde{S}$ and B agree at some times, but the time on $\tilde{S}$ is fixed while that on B is random. So what we want to apply is the law of large numbers. By the strong law of large numbers,

$$\lim_{n\to\infty} \frac{1}{n} |T_n^{(1)} - n| \to 0 \text{ as } n \to 0.$$

This implies that

$$\sup_{1\leq n\leq N} \frac{1}{N} |T_n^{(1)} - n| \to 0 \text{ as } N \to \infty.$$

Note that $(T_n^{(1)})_{n\geq 0} \sim (T_n^{(N)})_{n\geq 0}$, it follows for any $\delta > 0$,

$$\mathbb{P}\left(\sup_{1\leq n\leq N} \left|\frac{T_n^{(N)}}{N} - \frac{n}{N}\right| \geq \delta\right) \to 0 \text{ as } N \to \infty.$$

Using $(*)$ and continuity, for any $t \in [\frac{n}{N}, \frac{n+1}{N}]$, there exists $u \in [T_{n/N}^{(N)}, T_{(n+1)/N}^{(N)}]$ such that

$$\tilde{S}_t^{(N)} = B_u.$$

Note that if times are approximated well up to δ, then $|t - u| \leq \delta + \frac{1}{N}$.

Hence we have

$$\begin{aligned}\{\|\tilde{S} - B\|_\infty > \varepsilon\} \leq &\left\{\left|\tilde{T}_n^{(N)} - \frac{n}{N}\right| > \delta \text{ for some } n \leq N\right\} \\ &\cup \left\{|B_t - B_u| > \varepsilon \text{ for some } t \in [0,1], |t-u| < \delta + \frac{1}{N}\right\}.\end{aligned}$$

The first probability $\to 0$ as $n \to \infty$. For the second, we observe that $(B_t)_{T\in[0,1]}$ has uniformly continuous paths, so for $\varepsilon > 0$, we can find $\delta > 0$ such that the second probability is less than ε whenever $N > \frac{1}{\delta}$ (exercise!).

So $\tilde{S}^{(N)} \to B$ uniformly in probability, hence converges uniformly in distribution. □

6 Large deviations

So far, we have been interested in the average or "typical" behaviour of our processes. Now, we are interested in "extreme cases", i.e. events with small probability. In general, our objective is to show that these probabilities tend to zero very quickly.

Let $(X_n)_{n\geq 0}$ be a sequence of iid integrable random variables in $\mathbb{R}$ and mean value $\mathbb{E}X_1 = \bar{x}$ and finite variance σ^2. We let

$$S_n = X_1 + \cdots + X_n.$$

By the central limit theorem, we have

$$\mathbb{P}(S_n \geq n\bar{x} + \sqrt{n}\sigma a) \to \mathbb{P}(Z \geq a) \text{ as } n \to \infty,$$

where $Z \sim N(0, 1)$. This implies

$$\mathbb{P}(S_n \geq an) \to 0$$

for any $a > \bar{x}$. The question is then how fast does this go to zero?

There is a very useful lemma in the theory of sequences that tells us this vanishes exponentially quickly with n. Note that

$$\mathbb{P}(S_{m+n} \geq a(m + n)) \geq \mathbb{P}(S_m \geq am)\mathbb{P}(S_n \geq an).$$

So the sequence $\mathbb{P}(S_n \geq an)$ is super-multiplicative. Thus, the sequence

$$b_n = -\log \mathbb{P}(S_n \geq an)$$

is sub-additive.

Lemma (Fekete)**.** If b_n is a non-negative sub-additive sequence, then $\lim_n \frac{b_n}{n}$ exists.

This implies the rate of decrease is exponential. Can we do better than that, and point out exactly the rate of decrease?

For $\lambda \geq 0$, consider the *moment generating function*

$$M(\lambda) = \mathbb{E}e^{\lambda X_1}.$$

We set $\psi(\lambda) = \log M(\lambda)$, and the *Legendre transform* of ψ is

$$\psi^*(a) = \sup_{\lambda \geq 0}(a\lambda - \psi(\lambda)).$$

Note that these things may be infinite.

Theorem (Cramér's theorem)**.** For $a > \bar{x}$, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) = -\psi^*(a).$$

Note that we always have

$$\psi^*(a) = \sup_{\lambda \geq 0}(a\lambda - \psi(\lambda)) \geq -\psi(0) = 0.$$

Proof. We first prove an upper bound. For any λ, Markov tells us

$$\begin{aligned}\mathbb{P}(S_n \geq an) = \mathbb{P}(e^{\lambda S_n} \geq e^{\lambda an}) &\leq e^{-\lambda a_n}\mathbb{E}e^{\lambda S_n}\\ &= e^{-\lambda an}\prod_{i=1}^n \mathbb{E}e^{\lambda X_i} = e^{-\lambda an}M(\lambda)^n = e^{-n(\lambda a - \psi(\lambda))}.\end{aligned}$$

Since λ was arbitrary, we can pick λ to maximize $\lambda a - \psi(\lambda)$, and so by definition of $\psi^*(a)$, we have $\mathbb{P}(S_n \geq a_n) \leq e^{-n\psi^*(a)}$. So it follows that

$$\limsup \frac{1}{n}\log \mathbb{P}(S_n \geq a_n) \leq -\psi^*(a).$$

The lower bound is a bit more involved. One checks that by translating X_i by a, we may assume $a = 0$, and in particular, $\bar{x} < 0$.

So we want to prove that

$$\liminf_n \frac{1}{n}\log \mathbb{P}(S_n \geq 0) \geq \inf_{\lambda \geq 0} \psi(\lambda).$$

We consider cases:

- If $\mathbb{P}(X \leq 0) = 1$, then

  $$\mathbb{P}(S_n \geq 0) = \mathbb{P}(X_i = 0 \text{ for } i = 1, \dots n) = \mathbb{P}(X_1 = 0)^n.$$

  So in fact

  $$\liminf_n \frac{1}{n}\log \mathbb{P}(S_n \geq 0) = \log \mathbb{P}(X_1 = 0).$$

  But by monotone convergence, we have

  $$\mathbb{P}(X_1 = 0) = \lim_{\lambda \to \infty} \mathbb{E}e^{\lambda X_1}.$$

  So we are done.

- Consider the case $\mathbb{P}(X_1 > 0) > 0$, but $\mathbb{P}(X_1 \in [-K, K]) = 1$ for some K. The idea is to modify X_1 so that it has mean 0. For $\mu = \mu_{X_1}$, we define a new distribution by the density

  $$\frac{\mathrm{d}\mu^\theta}{\mathrm{d}\mu}(x) = \frac{e^{\theta x}}{M(\theta)}.$$

  We define

  $$g(\theta) = \int x \,\mathrm{d}\mu^\theta(x).$$

  We claim that g is continuous for $\theta \geq 0$. Indeed, by definition,

  $$g(\theta) = \frac{\int x e^{\theta x}\,\mathrm{d}\mu(x)}{\int e^{\theta x}\,\mathrm{d}\mu(x)},$$

  and both the numerator and denominator are continuous in θ by dominated convergence.

  Now observe that $g(0) = \bar{x}$, and

  $$\limsup_{\theta \to \infty} g(\theta) > 0.$$

So by the intermediate value theorem, we can find some θ_0 such that $g(\theta_0) = 0$.

Define $\mu_n^{\theta_0}$ to be the law of the sum of n iid random variables with law μ^{θ_0}. We have

$$\mathbb{P}(S_n \geq 0) \geq \mathbb{P}(S_n \in [0, \varepsilon n]) \geq \mathbb{E} e^{\theta_0(S_n - \varepsilon n)} \mathbf{1}_{S_n \in [0, \varepsilon n]},$$

using the fact that on the event $S_n \in [0, \varepsilon n]$, we have $e^{\theta_0(S_n - \varepsilon n)} \leq 1$. So we have

$$\mathbb{P}(S_n \geq 0) \geq M(\theta_0)^n e^{-\theta_0 \varepsilon n} \mu_n^{\theta_0}(\{S_n \in [0, \varepsilon n]\}).$$

By the central limit theorem, for each fixed ε, we know

$$\mu_n^{\theta_0}(\{S_n \in [0, \varepsilon n]\}) \to \frac{1}{2} \text{ as } n \to \infty.$$

So we can write

$$\liminf_n \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \geq \psi(\theta_0) - \theta_0 \varepsilon.$$

Then take the limit $\varepsilon \to 0$ to conclude the result.

– Finally, we drop the finiteness assumption, and only assume $\mathbb{P}(X_1 > 0) > 0$. We define ν to be the law of X_1 condition on the event $\{|X_1| \leq K\}$. Let ν_n be the law of the sum of n iid random variables with law ν. Define

$$\psi_K(\lambda) = \log \int_{-K}^{K} e^{\lambda x} \, \mathrm{d}\mu(x)$$
$$\psi^\nu(\lambda) = \log \int_{-\infty}^{\infty} e^{\lambda x} \, \mathrm{d}\nu(x) = \psi_K(\lambda) - \log \mu(\{|X| \leq K\}).$$

Note that for K large enough, $\int x \, \mathrm{d}\nu(x) < 0$. So we can use the previous case. By definition of ν, we have

$$\mu_n([0, \infty)) \geq \nu([0, \infty)) \mu(|X| \leq K)^n.$$

So we have

$$\begin{aligned}
\liminf_n \frac{1}{n} \log \mu([0, \infty)) &\geq \log \mu(|X| \leq K) + \liminf \log \nu_n([0, \infty)) \\
&\geq \log \mu(|X| \leq K) + \inf \psi^\nu(\lambda) \\
&= \inf_\lambda \psi_K(\lambda) \\
&= \psi_K^\lambda.
\end{aligned}$$

Since ψ_K increases as K increases to infinity, this increases to some $\mathcal{J}$ we have

$$\liminf_n \frac{1}{n} \log \mu_n([0, \infty)) \geq \mathcal{J}. \tag{$\dagger$}$$

Since $\psi_K(\lambda)$ are continuous, $\{\lambda : \psi_K(\lambda) \leq \mathcal{J}\}$ is non-empty, compact and nested in K. By Cantor's theorem, we can find

$$\lambda_0 \in \bigcap_K \{\lambda : \psi_K(\lambda) \leq \mathcal{J}\}.$$

So the RHS of ($\dagger$) satisfies

$$\mathcal{J} \geq \sup_K \psi_K(\lambda_0) = \psi(\lambda_0) \geq \inf_\lambda \psi(\lambda). \qquad \square$$

Index

$C([0,\infty),\mathbb{R})$, 29
$C([0,\infty),\mathbb{R})$, 29
$D([0,\infty),\mathbb{R})$, 29
$L^1(E)$, 6
L^p inequality, 21, 35
$M(\lambda)$, 57
$U[a,b,(x_n)$, 18
$U_n[a,b,(x_n)]$, 18
X_T, 15
Δ, 48
$\mathcal{B}$, 4
$\mathcal{B}(E)$, 4
$\mathcal{F}_T$, 16
$\mathcal{F}_\infty$, 31
σ-field, 4
σ-algebra, 4
    Borel, 4
    countably generated, 26
    product, 6
    trivial, 24
$m\mathcal{E}$, 4
$m\mathcal{E}^+$, 4

adapted process, 14
almost everywhere, 5
atom, 26

backwards filtration, 24
backwards martingale, 24
Blumenthal's 0-1 law, 45
Borel σ-algebra, 4
Brownian motion, 42
Brownian scaling, 44

cadlag function, 29
cadlag stochastic process, 29
characteristic function, 40
continuous stochastic process, 29
continuous time filtration, 31
continuous time martingale, 33
continuous time stochastic process, 29
continuous time sub-martingale, 33
continuous time super-martingale, 33
convergence in distribution, 37
countably generated σ-algebra, 26
Cramér's theorem, 57

Dirac δ measure, 37
discrete stochastic process, 14
distribution
    finite-dimensional, 30
distribution function, 38
domain, 48
dominated convergence theorem, 6, 11
Doob's L^p inequality, 21, 35
Doob's upcrossing lemma, 19
dyadic numbers, 30

events, 6

Fatou's lemma, 5, 11
Fekete's lemma, 57
filtration, 14
    backwards, 24
    continuous time, 31
    natural, 14
    right continuous, 33
finite-dimensional distribution, 30
Fubini's theorem, 6

harmonic function, 27, 48
hitting time, 32

increment, 42
integrable function, 6
integrable process, 14
integral, 5

Jensen's inequality, 11

Kolmogorov 0-1 law, 24
Kolmogorov's criterion, 30

Lévy's convergence theorem, 40
Laplacian operator, 48
law, 37
Legendre transform, 57

Markov chain, 27
Markov property, 44
martingale, 14
    continuous time, 33
martingale convergence theorem

L^1, 22, 36
L^p, 21, 36
almost sure, 18
maximal inequality, 20, 35
maximum principle, 50
mean value property, 48
measurable function, 4
measurable space, 4
measure, 4
product, 6
measure space, 4
moment generating function, 57

natural filtration, 14
neighbourhood recurrent, 52

optional stopping theorem, 16, 34

Poincaré cone condition, 50
point recurrent, 52
product σ-algebra, 6
product measure, 6
Prokhorov's theorem, 39

Radon–Nikodym derivative, 26
Radon–Nikodym theorem, 26
random fractal property, 44
random variables, 6
realization, 6
reflection principle, 46
regularization, 35
right continuous filtration, 33

simple function, 5
Skorokhod embedding theorem, 53
standard Brownian motion, 42
stochastic process, 29
adapted, 14
cadlag, 29
continuous, 29
discrete, 14
integrable, 14
stopped process, 15
stopping time, 15, 32
continuous time, 32
strong Markov property, 45
sub-martingale, 14
continuous time, 33
super-martingale, 14
continuous time, 33

tail σ-algebra, 24
tight probability measures, 39
Tonelli's theorem, 6
tower property, 11
transient, 52
transition matrix, 27
trivial σ-algebra, 24

upcrossing, 18
usual conditions, 33

version, 5, 35

weak convergence, 37
Wiener's theorem, 42